JUNE
1950
VOLUME 14
NUMBER 3

JOURNAL OF CONSULTING PSYCHOLOGY

AMERICAN PSYCHOLOGICAL ASSOCIATION

JOURNAL OF CONSULTING PSYCHOLOGY

EDITED BY LAURANCE F. SHAFFER

TEACHERS COLLEGE, COLUMBIA UNIVERSITY

ASSOCIATE EDITORS: EDGAR A. DOLL, *Devereux Schools, Devon, Pa.* WILLIAM A. HUNT, *Northwestern University.* E. LOWELL KELLY, *University of Michigan.* MORRIS KRUGMAN, *Board of Education, New York, N. Y.* BERTHA M. LUCKEY, *Board of Education, Cleveland, Ohio.* FRED MCKINNEY, *University of Missouri.* ANNE ROE, *New York, N. Y.* CARL R. ROGERS, *University of Chicago.* R. NEVITT SANFORD, *University of California.*

PUBLISHED BI-MONTHLY BY THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

THE JOURNAL OF CONSULTING PSYCHOLOGY is the clinical journal of the American Psychological Association. It is devoted primarily to the publication of original research relevant to psychological diagnosis, psychotherapy and counseling, personality, and the dynamics of behavior. Case studies, theoretical contributions, descriptions of clinical techniques, and discussions of the professional problems of clinical psychology also appear in its pages.

EDITORIAL OFFICE

Manuscripts submitted for publication, books and tests for review, and correspondence on editorial matters should be sent to the Editor, Laurance F. Shaffer, 525 West 120th St., New York 27, N. Y.

Information for contributors: The important characteristics of a good article are the significance of its subject matter, the demonstrated soundness of its conclusions, its clarity, and its brevity. Clarity is improved by careful outlining, discriminating use of headings, and a simple, straightforward style of writing. For brevity, authors should strive to convey every thought without waste of words, and should avoid repetitions of methods, data or findings.

Some of the mechanical requirements for manuscripts are: Type manuscripts with wide margins, on white bond paper. Double space between *each and every line*, including references, footnotes, matter to be set in smaller type, and tables. Use only two types of headings: centered headings, and underlined paragraph headings. Type each table, however small, on a separate sheet, with Arabic number and title. Draw all figures in India ink in final form for reproduction, with draftsman's lettering, but type captions for figures on separate sheets. The *References* must follow the exact form used in APA journals. For further information on the details of preparing manuscripts, see:

> ANDERSON, J. E., AND VALENTINE, W. L. The preparation of articles for publication in the journals of the American Psychological Association. *Psychol. Bull.*, 1944, 14, 345-376.

There is a lag of about nine to twelve months between the acceptance of an article and its publication. Authors may secure immediate publication, however, by arranging for articles to be printed as extra pages, increasing the number of pages that subscribers receive. The cost to the author of such early publication is approximately $11.00 per printed page.

Authors receive fifty reprints gratis, and may order additional quantities when returning proof. Authors are charged one-half of the extra cost of tabular material and cuts, and the full cost of changes made on the proof.

BUSINESS OFFICE

Communications concerning subscriptions, change of address, claims for the nonreceipt of a number, advertising, and other business matters should be sent to the American Psychological Association, Inc., 1515 Massachusetts Ave., N. W., Washington 5, D. C.

The Journal of Consulting Psychology is published bi-monthly in February, April, June, August, October, and December, the yearly volume comprising approximately 450 pages. The subscription per year is $5.00, foreign $5.50, single number $1.00.

Copyright 1950 by the American Psychological Association, Inc.

Printed and mailed by the Dentan Printing Co., 23 West Colorado Ave., Colorado Springs, Colo. and entered as second-class matter at the Post Office at Colorado Springs, Colo. Acceptance for mailing at special rate of postage provided for in section 538, Act of February 25, 1925, authorized May 27, 1940.

JOURNAL OF CONSULTING PSYCHOLOGY

June, 1950 Vol. 14, No. 3

CONTENTS

Configural Scoring: PAUL E. MEEHL - - - - - - - - - - - - - - - 165

The Reliability of the Wechsler-Bellevue Subtests and Scales: GORDON F. DERNER, MURRAY ABORN, AND AARON H. CANTER - - - - - - - - - - - - - - 172

The Diagnostic Significance of an Objective Measure of Intratest Scatter on the Wechsler-Bellevue Intelligence Scale: JULES D. HOLZBERG AND MAURICE A. DEANE - 180

Personality Problems of the Spinal Cord Injured: ALFRED D. MUELLER - - - - - 189

The Measurement of Intelligence of 102 Male Paraplegics: MORSE P. MANSON - - 193

The Effect of Focal Brain Lesions on Recall and on the Production of Rotations in the Bender Gestalt Test: LEO J. HANVIK AND A. LLOYD ANDERSEN - - - - - 197

A Basis for Integration Rather Than Fragmentation in Psychotherapy: REX MADISON COLLIER - 199

Ego Involvement and the Process of Disengagement: IJA N. KORNER - - - - - 206

An Investigation of Nondirective Group Therapy with Students in Academic Difficulty: WILLIAM D. SHELDON AND THEODORE LANDSMAN - - - - - - - - - - 210

The Implications of Certain Assumptions Involved in the Use of the TAT: J. R. WITTENBORN - 216

Objective Rorschach Response Patterns Differentiating Anxiety Reactions from Obsessive-Compulsive Reactions: SOLIS L. KATES - - - - - - - - - - - - 226

The Susceptibility to Falsification of the Rorschach Psychodiagnostic Technique: ABRAHAM L. CARP AND ARTHUR R. SHAVZIN - - - - - - - - - - - 230

A Note on the Validity of the Leiter International Performance Scale: TRENT E. BESSENT - 234

New Books and Tests - 235

CONFIGURAL SCORING[1,2]

PAUL E. MEEHL
UNIVERSITY OF MINNESOTA

ONE of the most important words in the vocabulary of clinicians is the word "pattern." We speak of ourselves as thinking in terms of totalities, organizations, configurations; and we congratulate ourselves for being able to look upon case material in this patterned, non-atomistic way. The question as to how much we really do this, and the further question of how well it pays off in terms of predictive efficiency, is a very complicated one on which neither theoretical nor adequate empirical data as yet exist. But I am sure everyone, even the most "hardheaded," who is at all acquainted with clinical work will at least agree that there are situations in which data are integrated in a "patterned" manner. The word *patterning* is ambiguous, and I have tried elsewhere to give a formal (mathematical) definition of patterning of variables for the continuous case [16]. One obvious, simple instance of patterning is the case in which a datum is taken as evidence *for* a certain hypothesis about the patient when seen in a certain context of other evidence; whereas the same datum would be taken as arguing *against* the very same hypothesis with the context different. For example, an objective history of extremely wide and vigorous athletic participation in high school would, statistically speaking, argue in favor of masculinity in the male. If we find such a history in a male aged 35, without heterosexual experience, living with his mother, and "sponsoring" boys' clubs, we begin to think of the latent homoerotic component as being strong; and the exaggerated athletic activity can be viewed as defense. So viewed, it is actually part of the evidential picture in favor of the hypothesis that the patient has a problem in the area of sexual role. We are all aware of the dangers involved in this kind of thinking, which has at times a certain "heads I win, tails you lose" flavor; but most of us would say that a person who cannot stand this kind of ambiguity in the evidence ought to get out of clinical psychology, because it is simply the case that such things as reaction formation *do* occur, and if you want to deal with human motivation as a subject matter you have to take the consequences.

A frequently voiced objection to structured personality tests is couched in terms of this configural character of clinical material. It is pointed out, quite justly, that the score on a test variable such as the Multiphasic *Pd* is obtained by simply adding the number of items to get a total, with no attention being paid to the "meaning" and "pattern" of the responses. While recent clinical use of instruments such as the Multiphasic has taken for granted the emphasis upon profile patterns rather than an atomistic emphasis upon single scales [1, 3, 5, 6, 7, 8, 11, 15, 18, 19, 21, 22], this refers to patterning of the scored *variables* and does not meet the problem at the level of determining the single score itself, i.e., at the item level.

An interesting paradox which I developed recently during a preliminary oral examination seems to me relevant in this connection. I take the liberty of calling this a paradox because practically all the psychologists I have since asked about it give the same answer as the candidate, which is the wrong one. The question is this: Consider two dichotomously scored test items, which we wish to use in predicting a dichotomous criterion. Let us ignore sampling errors and confine the discussion entirely to parameters. Suppose that the dichotomous criterion is, say, "schizophrenic" versus "normal." For the present purpose the question as to whether some continuous variable or set of

[1]Written while on a (summer session) research appointment granted by the University of Minnesota Graduate School.

[2]I am indebted to Dr. Kenneth E. Clark for stimulating discussion and clarification of several points treated in this paper.

variables actually underlies such a clinical dichotomy is irrelevant.

Suppose now that (in the supply) *each* of the two items has exactly 50 per cent difficulty within *each* category, so that half of the schizophrenics and half of the normals answer each of the items "true" and half of each group answer each item "false." Under these circumstances any of the usual methods of item analysis (phi-coefficient, significance of difference of proportions, tetrachoric *r*, and so forth) will show both items to have "zero validity" for the criterion, and they will be eliminated. Under such conditions, is it possible to predict the criterion solely upon the basis of the response to these two items, and if so, how well could it theoretically be predicted? Almost everyone who is asked this question has the initial tendency to say that the criterion could not be predicted at all; and most people finally come out with this answer even after some paranoid ruminations. It is, therefore, interesting to consider an extreme case.

TABLE 1
INTERITEM CORRELATION AMONG NORMALS

		Item No. 2 T	F	
Item No. 1	T	50	0	50
	F	0	50	50
		50	50	100

$\phi_1 = +1.00$

TABLE 2
INTERITEM CORRELATION AMONG SCHIZOPHRENICS

		Item No. 2 T	F	
Item No. 1	T	0	50	50
	F	50	0	50
		50	50	100

$\phi_2 = -1.00$

TABLE 3
VALIDITY OF ITEM-PAIR WHEN SCORED CONFIGURALLY

		Normal	Schiz.	
Response Pattern	TT or FF	100	0	100
	TF or FT	0	100	100
		100	100	200

$\phi_f = +1.00$
Validity of each item = 0.00

Table 1 shows the correlation between the two items in the first criterion category and tells us that all normal persons answer the two items in the "same way," although they do not have any tendency to answer them either true or false. Table 2 shows us that schizophrenics, while also having no tendency to answer either item true or false, are consistent in another way, since they always give "opposite" answers to the two items. However, each item, considered singly, has no discriminating power. Table 3 shows the predictive efficiency of scoring these items *configurally*. That is, we consider the response to the two items simultaneously on the scoring key rather than one at a time. Any person who answers the two items in the *same way* (either both true or both false), is predicted to be a normal; anyone who answers the items in the *opposite* way, whether in the true-false or false-true pattern, is considered to be a schizophrenic. In this hypothetical ideal case, we find a perfect predictability of the dichotomous criterion on the basis of the two items, even though each item has zero validity. It is obvious that this represents simply the extreme case of situations which can exist in lesser degrees.

The general condition for this kind of pairwise scoring to create joint validity when each item singly has zero validity is that the *interitem* correlation should be different in the two categories of the criterion. In the ideal case just treated, this disparity between the interitem correlations in the two categories of the criterion is at a maximum, since the interitem phi-coefficient among the schizophrenics is —1.00, and the corresponding interitem phi-coefficient among the normals is +1.00. In a second artificial example, the discrepancy is considerably less although the interitem phi-coefficient does differ in sign for the two portions of the population which we wish to separate.

Table 4 shows that the interitem correlation among normals is .40, whereas the inter-

TABLE 4
INTERITEM CORRELATION AMONG NORMALS

		Item No. 2 T	F	
Item No. 1	T	35	15	50
	F	15	35	50
		50	50	100

$\phi_1 = +.40$

TABLE 5
INTERITEM CORRELATION AMONG SCHIZOPHRENICS

		Item No. 2 T	Item No. 2 F	
Item No. 1	T	15	35	50
	F	35	15	50
		50	50	100

$\phi_2 = -.40$

TABLE 6
VALIDITY OF ITEM-PAIR WHEN SCORED CONFIGURALLY

		Normal	Schiz.	
Response Pattern	TT or FF	70	30	100
	TF or FT	30	70	100
		100	100	200

$\phi_f = +.40$
Validity of each item = 0.00

item correlation among schizophrenics (Table 5) is of the same absolute value but opposite in sign. Under these conditions, scoring the items configurally so that again a true-true or false-false pattern of response is considered normal and a true-false or false-true pattern is considered schizophrenic, we get the frequencies of Table 6, showing that the item-pair, scored configurally, has a validity coefficient of .40. The validity coefficient against the dichotomous criterion obtained from configural scoring of a pair of items when the interitem phi-coefficients are located symmetrically about zero is always the same as the absolute value of the phi-coefficients. Thus, two items each of which has zero validity but which have a correlation of —.60 in one class and +.60 in the other, would yield when scored configurally a validity coefficient of .60 in separation of the criterion classes.

Consider next the case in which, although still opposite in sign, the interitem phi-coefficients are not symmetrical about zero.

TABLE 7
INTERITEM CORRELATION AMONG NORMALS

		Item No. 2 T	Item No. 2 F	
Item No. 1	T	37	13	50
	F	13	37	50
		50	50	100

$\phi_1 = +.48$

TABLE 8
INTERITEM CORRELATION AMONG SCHIZOPHRENICS

		Item No. 2 T	Item No. 2 F	
Item No. 1	T	22	28	50
	F	28	22	50
		50	50	100

$\phi_2 = -.12$

TABLE 9
VALIDITY OF ITEM-PAIR WHEN SCORED CONFIGURALLY

		Normal	Schiz.	
Response Pattern	TT or FF	74	44	118
	TF or FT	26	56	82
		100	100	200

$\phi_f = +.31$
Validity of each item = 0.00

Still holding the single item validities to zero, in Table 7 the interitem correlation among normals is .48, while the corresponding interitem correlation among schizophrenics is —.12. Calling the T-T or F-F configurations normal and the T-F or F-T configurations schizophrenic, we see the validity for configural scoring is at .31 (Table 9). The algebraic difference of the phi-coefficients in the two categories of the population is .60, so that if such a difference ($\phi_1 - \phi_2$) had been symmetrical about zero, i.e., an interitem phi of .30 in the normals and —.30 in the schizophrenics, we would have a configural validity of exactly .30. So we see that a considerable asymmetry does not markedly affect the magnitude of validity obtained through configural scoring when the numbers are in this general range.

Next consider a case which is intuitively somewhat less satisfying but still shows how the configural scoring can generate validity. Tables 10 and 11 represent the situation of an item-pair which has a very high positive association among normals and a positive but considerably lower association among schizophrenics. In this instance, for both portions of the population the "expected" (modal) response pattern is one of agreement. That is, we expect the schizophrenics as well as the normals to answer both items in the same way, even though we have no expectations as to *which way* they will answer them. Under such circumstances if we score a T-F or F-T pattern as schizophrenic, it might seem somewhat unreasonable, as this is not

what most of the schizophrenics do. But the alternative leaves us without any basis for making predictions. In Table 12 we see the results of scoring configurally by calling all T-T and F-F response patterns normal, and all T-F and F-T response patterns schizophrenic. This yields a configural validity of .37 which, it will be noted, is .07 higher than that obtained with a similar algebraic difference of the two phi-coefficients when they are symmetrical about zero.

TABLE 10
INTERITEM CORRELATION AMONG NORMALS

		Item No. 2 T	F	
Item No. 1	T	47	3	50
	F	3	47	50
		50	50	100

$\phi_1 = +.88$

TABLE 11
INTERITEM CORRELATION AMONG SCHIZOPHRENICS

		Item No. 2 T	F	
Item No. 1	T	32	18	50
	F	18	32	50
		50	50	100

$\phi_2 = +.28$

TABLE 12
VALIDITY OF ITEM-PAIR WHEN SCORED CONFIGURALLY

		Normal	Schiz.	
Response Pattern	TT or FF	94	64	158
	TF or FT	6	36	42
		100	100	200

$\phi_j = +.37$
Validity of each item = 0.00

Since we are only now in the process of collecting some "real" material on MMPI test items, it has not seemed profitable to enter into the mathematical properties of such systems in any detail. I have, however, plotted the configural validity coefficient as a function of the algebraic difference of the interitem correlations in the two criterion categories, plotting a separate such curve for various values of the point about which the phi-coefficients are symmetrical. For points of symmetry as high as .60, the relationship between configural validity and the difference $(\phi_1 - \phi_2)$ is very close to linear; and as we saw for symmetry about zero is precisely so. The functions, however, become slightly positively accelerated for points of symmetry other than zero, so that the estimated configural validity obtained by taking the value

$$\frac{\phi_1 - \phi_2}{2}$$

is an underestimate. At the extremes this introduces a considerable error. For example, consider a pair of items for which the phi-coefficients in the two categories to be discriminated are 1.00 and .80. The difference $(\phi_1 - \phi_2)$ is .20, which leads to an estimate of about .10 for the configural validity. However, when the point of symmetry is so far from zero as .90, the actual configural validity is approximately .23.

I have not been able to derive an exact formula for the configural validity expressed solely as a function of the two phi-coefficients involved, although it is of course easy to get an expression for a joint (configural) validity in terms of the proportions in the eight cells of the two fourfold tables. If the technique should turn out to be of practical value, graphical methods would be adequate for purposes of selecting potentially valuable item pairs. The sampling characteristics of the joint phi-coefficients have not been investigated.

If items which behave in this way can be found in the field of personality and interest measurement, it seems to me that such a technique would have considerable practical importance. In the first place we have the straightforward argument that if validity can be achieved with items that have no validity considered singly, it may be possible to develop instruments of a structured type to perform functions which are not generally accessible to structured tests. A second important consideration is in the matter of subtlety. While advances have been made in recent years in the direction of an increase in subtlety of structured personality measuring instruments and in detection of distortions [4, 6, 9, 12, 13, 17, 24], it must be admitted that very serious problems in this area still remain [2, 4, 10, 11, 20]. It is probable that these problems will be much more difficult to solve with respect to certain dimensions, e.g., paranoid components of personality, than with respect to others, e.g., som-

atization components. For example, the *Pa* scale of the MMPI is a relatively weak scale in the sense that a very sizable subgroup of clinically grossly paranoid individuals are able somehow to evade a high score without giving themselves away on the validity indicators. It would seem that the possibilities for subtlety when configural scoring is made use of are greatly enhanced since it would require much more sophistication to know in what way certain verbal responses would tend to "hang together" in variously defined groups of people than it would to know, in general, the kind of response which is "healthy" as opposed to "maladjusted."

If such items can be found in a pool such as that of MMPI, there might also be considerable theoretical significance in that it would leave us with a problem of interpretation from the factorial standpoint. For instance, at the present time, the theoretical conceptualization of the relation between so-called "obvious" and "subtle" items is very obscure. Wiener [22] has interpreted his findings on subtle-obvious relations in successful and unsuccessful salesmen by assuming that the subtle items of such scales as *Pa* reflect the paranoid tendency faithfully, but that those who show a very high subtle score contrasted with their score on the obvious component have learned to recognize their own abnormality and to avoid the more concrete overt expression of such tendencies. For this reason persons with subtle *Pa* higher than obvious seem by certain external criteria to be better adjusted. I am certainly not in a position to refute Wiener's hypothesis on any data available to me; but I should like to emphasize that there is an alternative interpretation which is as meaningful psychologically and would lead to the same psychometric consequences. Consider such an item as: "Some people are so bossy that I feel like doing the opposite of what they request even though I know they are right" which is scored for *Pa* if answered *False*. This is an excellent example of *Pa* subtle, both in terms of its obvious content which is superficially *non*paranoid, and also in terms of its statistical property, namely, that the answer scored as paranoid is the answer given by the majority of the unselected normals of the standardization population. As has been pointed out in previous papers [17, 23] this item appears also on the Guilford-Martin Personnel Inventory where the response *False* is scored as *non*paranoid (i.e., opposite to the MMPI scoring, derived empirically on paranoid psychotics). Consider three groups of individuals who are not qualitatively distinct but who are clustered in three general loci on a paranoid dimension of some sort. Since this is a "zero" item, the "normal nonparanoid" individual in the general population will answer the item in the scored direction. That is, most normal, nonsuspicious people will *not* report that "Some people are so bossy that I feel like doing the opposite of what they request even though I know they are right." The preponderance of such nonsuspicious people in the general population sample accounts for the fact that this item is scored with a "0" on the *Pa* scoring template. Consider now a second group of people, also well within the normal (nondiagnosible, nonpsychotic) population, but who are somewhat more suspicious, stubborn, and hostile toward mankind than the previous group. One would anticipate that such people would tend to give the "obvious" paranoid response to this item, i.e., they would tend to answer it *True*. Item derivations operating within the nonpsychotic range whether by an external criterion or by methods of internal consistency such as were used by Guilford, would assign the response "True" a scoring in the paranoid, suspicious, uncooperative direction—a scoring which, *operating within this range*, may very well be the valid one. Finally, let us consider a full-blown diagnosable paranoid psychotic who answers such obvious items as "Someone has it in for me" (*True*), "I am sure I am being talked about" (*True*), "I believe I am being plotted against" (*True*), and so on. This paranoid psychotic with reality-distortions of delusional proportions has set up in addition the pseudo-objectivity which is a well-known facet of many of these patients as we see them clinically. They sometimes surprise us with the cautiousness and apparent intellectual honesty with which they consider alternatives to their formulations. As is true of many writers on controversial subjects, they raise the opponent's objections and proceed to deal with them—the only trouble being that they do not raise the right ones. Whether this almost pedantic pseudo-objectivity of fairly intact paranoid patients is a defense erected following

earlier rebuff of their ideas by others, or a reaction against the suspiciousness which can afford to be spread widely over the area of mentation provided it does not encroach upon the specific delusional content, it is not important here to discuss. The main point is that such a clinically diagnosable paranoid patient begins to reverse the trend which was detectable as we moved toward the more suspicious and hostile end of the normal range. So that this patient (unlike the "normal paranoid" in the general population) gives the same response to the *Pa* subtle item as do well-adjusted, nonparanoid normals, although for very different reasons. This kind of thinking leads one to wonder whether the correlational properties of test items are not to be separately studied with the clinical picture and psychodynamics constantly in mind, at different positions on the continuum to be discriminated. It might turn out that people who give the response "Someone has it in for me" (*False*), and also the response "Some people are so bossy, etc." (*False*) are normal, well-adjusted, nonsuspcious people. Those who give the response "Someone has it in for me" (*True*) and "Some people are so bossy, etc." (*True*) are suspicious, hostile individuals in the nonclinical range, with also some tendencies to plus-getting on structured tests [17]. And finally those (very few) persons who give the response "Someone has it in for me" (*True*) but on the other hand *also* give the response "Some people are so bossy, etc." (*False*) are full-blown diagnosable paranoid psychotics who have moved in their illness to the point where they have developed the typical overmeticulous rigid pseudo- objectivity of the sick paranoid. It is obvious that the usual method of adding points of the *Pa* scale and similar instruments would not do justice to the complexity of these relationships if they exist. It remains to be seen whether this complexity, which everybody knows occurs in the interview situation and in the psychopathology of the patients, manifests itself sufficiently clearly to be useful in the verbal responses to the materials of a structural inventory.

The mechanics of "searching" for such configural scored item-pairs present practical difficulties. With a relatively small pool of items to begin with it would be feasible to set up the material on punch-cards and compute all possible interitem phi-coefficients separately for each of the two groups to be discriminated. Arranging the phi-coefficients in parallel columns one could simply pick those item-pairs which showed a large discrepancy (algebraic difference) between ϕ_1 and ϕ_2. With any appreciable number of items, this procedure is too laborious, particularly when a priori considerations do not suffice to select "likely" pairs. Further, the aim of subtlety in configural scoring might be defeated by the selection of item pairs on theoretical considerations. Short cut methods of approaching the exact "empirical" procedure will be developed if the pilot studies in progress indicate sufficient advantages in the technique to warrant its development.

The pair-wise configural scoring is the simplest of a whole series of configural scoring procedures which could eventually amount to the scoring of fairly large numbers of items simultaneously. For example, it might be that we give the patient one point for hysteroid tendencies if he answers item No. 37 *false, on condition that* he has answered items No. 46 and No. 52 oppositely to each other. In other words, considering these three items in numerical order, the configurations F-T-F and F-F-T would be scored for the trait in question; whereas the configurations T-T-T, T-T-F, T-F-T, T-F-F, F-T-T, F-F-F would receive no credit. Here again there are practical limitations upon the cumbersomeness of scoring, although the obvious ease and objectivity of scoring in structured tests is such that if it is necessary to complicate it considerably in order to gain some of the advantages of projective techniques it would be clinically justified.

I have treated only the special case where the items considered singly have zero validity. The question as to how much it would pay to employ a configural scoring of item pairs, triads, or *n*-ads when the component members already contain considerable validity is a complicated problem which I shall leave to mathematically abler workers. The aim of the present note is simply to bring this kind of scoring, for which I propose the designation *configural,* to the attention of clinicians and test-builders so that its possibilities may be explored in a diversity of testing areas.

Received July 27, 1949.

References

1. Abramson, H. A. The effect of alcohol on the personality inventory. *Anal. N.Y. Acad. Sci.*, 1946, **46**, 535-557.
2. Benton, A. L. The MMPI in clinical practice. *J. nerv. ment. Dis.*, 1945, **102**, 416-420.
3. Brown, M. A. Alcoholic profiles on the MMPI. Paper read at Midwest. Psychol. Ass., Chicago, 1949.
4. Cofer, C. N., Chance, J., and Judson, A. J. A study of malingering on the MMPI. *J. Psychol.*, 1949, **27**, 491-499.
5. Gough, H. G. Diagnostic patterns on the MMPI. *J. clin. Psychol.*, 1946, **2**, 23-37.
6. Gough, H. G. Simulated patterns on the MMPI. *J. abnorm. soc. Psychol.*, 1947, **42**, 215-225.
7. Guthrie, G. M. An evaluation of six MMPI diagnostic patterns. Paper read at Midwest. Psychol. Assn., Chicago, 1949.
8. Hathaway, S. R. A coding system for MMPI profiles. *J. consult. Psychol.*, 1947, **11**, 334-337.
9. Hovey, R. B. Detection of circumvention on MMPI. *J. clin. Psychol.*, 1943, **4**, 97.
10. Hunt, H. F. The effect of deliberate deception on MMPI performance. *J. consult. Psychol.*, 1948, **12**, 396-402.
11. Hunt, H. F., Carp, A., Cass, W. A., Winder, C. L., and Kantor, R. E. A study of the differential diagnostic efficiency of the MMPI. *J. consult. Psychol.*, 1948, **12**, 331-336.
12. Jurgensen, C. E Report on the "classification inventory," a personality test for industrial use. *J. appl. Psychol.*, 1944, **28**, 445-460.
13. McKinley, J. C., Hathaway, S. R., and Meehl, P. E. The MMPI: VI. The K scale. *J. consult. Psychol.*, 1948, **12**, 20-31.
14. Meehl, P. E. The dynamics of structured personality tests. *J. clin. Psychol.*, 1945, **1**, 296-303.
15. Meehl, P. E. Profile analysis of the MMPI in differential diagnosis. *J. appl. Psychol.*, 1946, **30**, 517-524.
16. Meehl, P. E. Clinical and actuarial methods of prediction.
17. Meehl, P. E. and Hathaway S. R. The K-factor as a suppressor variable in MMPI. *J. appl. Psychol.*, 1946, **30**, 525-564.
18. Modlin, H. C. A study of the MMPI in clinical practice, with notes on the Cornell Index. *Amer. J. Psychiat.*, 1947, **104**, 758-769.
19. Schmidt, H. O. Test profiles as a diagnostic aid: the MMPI. *J. appl. Psychol.*, 1945, **29**, 115-131.
20. Schmidt, H. O. Note on the MMPI: the K-factor. *J. consult. Psychol.*, 1948, **12**, 337-342.
21. Schneck, J. M. The double-spike pattern on the MMPI. *Amer. J. Psychiat.*, 1948, **104**, 443-445.
22. Welsh, G. S. An extension of Hathaway's MMPI coding system. *J. consult. Psychol.*, 1948, **12**, 343-344.
23. Wesley, E. Correlations between the Guilford-Martin personality factors O-Ag-Co and the MMPI at the college level. Unpublished M.A. thesis, Univ. of Minnesota, 1945.
24. Wiener, D. N. Subtle and obvious keys for the MMPI. *J. consult. Psychol.*, 1948, **12**, 164-170.
25. Wiener, D. N. Selecting salesmen with subtle-obvious keys for the MMPI. *Amer. Psychologist*, 1948, **3**, 364.

THE RELIABILITY OF THE WECHSLER-BELLEVUE SUBTESTS AND SCALES

GORDON F. DERNER, MURRAY ABORN, AND AARON H. CANTER
TEACHERS COLLEGE, COLUMBIA UNIVERSITY

SINCE its publication in 1939, the Wechsler-Bellevue Intelligence Scale [16] has received extensive clinical use, and many papers concerning it have appeared in the literature. Considerable emphasis has been placed on problems of psychiatric diagnosis, on attempts to prepare shorter scales, and on comparisons with other tests. The literature on the reliability of the several scales and their component subtests is, on the other hand, decidedly limited. Except for the studies reported by Wechsler [16] as part of the standardization procedure and Canter's [1] control population in his study of multiple sclerosis, the few reliability studies undertaken [7, 13, 14] have used neurotic or psychotic populations.

THE PROBLEM OF INTELLIGENCE TEST RELIABILITY

The commonly acceptable methods of determining test reliability are the split-half, alternate-form, and test-retest techniques. The serious limitation of the split-half technique is its failure to adequately indicate the dependability of the instrument over a period of time because the results show only the consistency of differentiation between individuals for a single administration. This method also requires a lengthy test, since dividing the test in half may shorten it to the extent that the estimate of reliability becomes excessively subject to chance fluctuations. While the Spearman-Brown prophecy formula can be used to estimate what the reliability coefficient would be if each half of the test were of double length, it can be employed only when the two halves demonstrate equal variability.

The method of determining reliability by the use of alternate forms demands that their equivalency be established upon the same sample. Forms are usually considered equivalent if they yield similar means, standard deviations, and distributions. The thinking here, however, becomes circular in that the tests must be reliable to be equivalent and equivalent to be reliable.

The method for investigating reliability by test-retest determines if a person's relative position in a sample remains the same over a period of time. Practice effect may cause the score on the retest to be higher than on the first administration but the consistency of practice effect can be noted. When short periods of time occur between retests familiarity with the testing procedures and the recall of responses given on the first administration may yield a somewhat spuriously high correlation. This problem can be dealt with by making retest measurements at varying time intervals. Inasmuch as interest is usually in the stability of an individual's score over varying periods of time, the test-retest technique should be the reliability method of first choice.

SPLIT-HALF RELIABILITY STUDIES

Wechsler [16, p. 133] used a variation of the split-half technique in his standardization. He chose four subtests—Digit Span, Picture Completion, Information, and Block Design—and correlated them with four other subtests —Arithmetic, Comprehension, Picture Arrangement, and Digit Symbol. The correlation for 355 subjects, ages 20-34, was .90 after correction for attenuation. The selection of subtests, however, seems to have been made to achieve maximal correlations. Subtests such as Object Assembly, whose correlations with other subtests of the scales were low, were omitted. A comparison of the Verbal IQ's and Performance IQ's for this same population yielded a correlation coefficient of .83 after correction for attenuation.

Actually, these are not reliability studies at all. The most valid inference to be made from

Wechsler's report is that the two chosen groups of subtests measure some common factor to a high degree. His correlations do not demonstrate the *consistency* with which they do so. The correlation between the Verbal and Performance IQ's may similarly be taken as the degree to which they measure a common factor, not as a measure of their dependability.

If Wechsler's definition of reliability were to be used, the several studies of abbreviated Wechsler-Bellevue scales [2, 4, 6, 8, 9, 10, 12, 15] would be pertinent. These studies do not purport to be reliability investigations, but attempt to select subtests of the scales which can be combined to measure intelligence to the same degree as the total scales.

A satisfactory split-half reliability study with the Wechsler-Bellevue is impractical for several reasons. There are too few items in each subtest to permit an adequate division into two halves. The heterogeneity of the subtests requires that the scale be divided by items rather than by subtests, if the concept of reliability is to have its usual meaning.

ALTERNATE FORM RELIABILITY STUDIES

There is no true alternate form of the Wechsler-Bellevue Scales although there is a second form [17] which Wechsler proposes as an alternate. The information Wechsler gives concerning its standardization is very limited, but the usual procedure of using the same population for the standardization of the alternate as for the original test was not followed. Unless the means, standard deviations, and distributions obtained for both forms are shown to be equal on the same population, a second form of a test cannot be considered an alternate. Because Form II does not meet these criteria, the findings of investigations using both forms are not acceptable as evidence of reliability but serve only as a comparison between two similarly conceived but distinct tests.

Gibby [5] tested 32 psychoneurotic patients with both forms of the Wechsler-Bellevue. The ages of these subjects ranged from 20 to 45 but 75 per cent were between 20 and 30. He gave both tests at one administration and altered his presentation of forms. Comprehension and Arithmetic were found to be significantly more difficult on Form II than on Form I. The correlation between Verbal IQ's and Performance IQ's was moderate, and that between the Full Scale IQ's high. The subtest and scale correlations are reported in this article in Table 5, for the sake of completeness, although they are not indices of reliability in the usual sense.

TEST-RETEST RELIABILITY STUDIES

Test-retest studies not only provide some basis for determining the expected consistency of an instrument but permit an estimate of practice effect. In Wechsler's report [16, p. 133] on reliability, the only genuinely pertinent study is the correlation between retests on 32 children ages 10-13, and 20 adults ages 20-34. The rank-difference method of correlation rather than the product-moment method was used. *Rho* was .94 for each group. The groups were very small, however, and there was no temporal control of practice effect. The individuals were retested at intervals which varied from one month to one year. There is no indication as to the range of IQ's represented nor as to the "normalness" of the subjects. Further, no report on the subtest reliabilities is included.

Rabin [13, 14] reported two studies on test-retest with the Wechsler-Bellevue. In one [14], he gives the test-retest correlation on 60 state hospitalized subjects of a wide age range and with miscellaneous diagnoses. The subjects were retested at intervals varying from 2 months to 27.8 months with a mean interval of 13.45 months. The correlations are reported in this paper in Table 5.

A second study by Rabin [13] compared test-retest results of 30 schizophrenics with 30 nonschizophrenic state hospitalized subjects with miscellaneous diagnoses including manic-depressive psychosis and psychoneurosis. The mean age of the schizophrenics was 28.1 years, as compared with a mean of 38.0 for the nonschizophrenics. The interval between retests ranged from 1 to 35 months with a mean interval of 13.07 months. The very marked differences between the results on schizophrenics and nonschizophrenics can be noted by reference to Table 5. The correlations between test-retest results for the mentally ill is lowest for schizophrenics. There are such considerable fluctuations that the results of the test for schizophrenics seem almost meaningless. There are, however, marked discrepancies between the findings on the schizophrenic population

in each study. Correlations are much higher for nonschizophrenics.

Hamister's findings [7], however, do not confirm Rabin. His study found no significant differences in subtest and scale reliabilities between a group of 34 schizophrenics, and a combined neuropsychiatric population of 53 subjects including the 34 schizophrenics. The similarity between the reliability coefficients obtained with these two groups is largely spurious, however, since there is the great overlapping of subjects between the two groups. The 34 schizophrenics had a mean age of 29.19 and the combined neuropsychiatric subjects had a mean age of 29.26. The age range of the combined group was from 18 through 40. Thirty-three patients were retested after a one-week interval and 20 were retested after a one-month interval. Because there were no significant differences between the means of the one-week and one-month group they were combined in the computation of correlation coefficients. The subtest and scale correlations are reported in Table 5.

Canter's investigation [1] of direct and indirect measures of psychological deficit included a test-retest study with the Wechsler-Bellevue on an experimental group of 47 patients with multiple sclerosis and a control group of 38 normal subjects. The age range of the sclerotics was from 20.5 to 47.2 with a mean age of 32.09. The age range of the normals was from 21.5 to 49.6 with a mean age of 31.24. All subjects were retested after a six-month interval so that temporal practice effect was held constant. His study was the first to make available any sizable, controlled sample for subtest reliability with a normal population. The findings on Canter's normal population have been incorporated into the present study. The correlations for the sclerotics are reported in Table 5.

PROBLEM

It is evident from a survey of the literature that the most fruitful study of reliability is a test-retest experiment with a large sample of normal subjects. No such study has been reported previously. A study of subtest reliability is especially important because of the use of deviant subtest scores as pathognomonics. Studies of the reliabilities of the various subtests that have been reported are for populations of neuropsychiatric subjects only. Finally, the effect on the reliability of varying lengths of time before retesting can give an indication of the stability of the test. The hypotheses raised are:

1. The Wechsler-Bellevue Scales have sufficiently high reliability to be considered diagnostically dependable.
2. The Wechsler-Bellevue subtests have sufficiently high reliability to be used for differential diagnosis by the deviant-score-sign technique.
3. The Wechsler-Bellevue Scales and subtests show no significant differences in reliability which are related to length of time before retesting.

PROCEDURE

Three groups of subjects were tested and retested for this study. Thirty-eight subjects were retested with a six-month interval between tests. This interval was within one week of the actual six-month date. Sixty subjects were retested with a four-week interval between tests. This interval was within one day of the actual four-week date. Sixty subjects were retested with a one-week interval between tests. This interval was on the exact one-week date. The six-month retest group ranged in age from 21.5 to 49.6 with a mean age of 31.24. The four-week and one-week groups fall in the age range 20 to 34, which was the age range of the group Wechsler selected for the determination of his weighted scores. The mean age for the four-week group was 25.80. The mean age for the one week-group was 25.46. All subjects were white as in the standardization population. The education-

TABLE 1
EDUCATIONAL LEVEL OF SUBJECTS IN PRESENT STUDY AND IN WECHSLER'S STANDARDIZATION POPULATION

Educational Category	Present Study %	Wechsler %
College graduates	10.00	5.10
Some college work	14.17	3.77
High school graduate only	60.83	10.81
Some high school work	11.67	18.76
Elementary school graduate only	3.33	28.85
Some elementary school work	0.00	30.17
Illiterate	0.00	2.55
TOTAL	100.00	100.01

TABLE 2

THE MEANS AND STANDARD DEVIATIONS FOR THE THREE GROUPS OF NORMAL SUBJECTS AND THE DIFFERENCES BETWEEN THEM OBTAINED IN THE ONE-WEEK, FOUR-WEEK AND SIX-MONTH RETESTS

Subtests	One-Week ($N = 60$)					Four-Week ($N = 60$)					Six-Month ($N = 38$)				
	M_1	M_2	D	SD_1	SD_2	d	M_1	M_2	D	SD_1	SD_2	d	M_1	M_2	D	SD_1	SD_2	d
Information	13.5	14.1	+0.6	1.9	1.9	0.0	12.9	13.7	+0.8	2.1	1.6	−0.5	12.1	12.7	+0.6	1.8	1.7	−0.1
Comprehension	12.5	12.9	+0.4	2.2	2.4	+0.2	12.5	12.8	+0.3	2.5	2.5	0.0	10.8	11.8	+1.0	2.2	2.5	+0.3
Digit Span	10.9	12.0	+1.1	2.7	3.1	+0.4	10.5	11.8	+1.3	2.9	2.9	0.0	9.9	10.8	+0.9	2.8	3.1	+0.3
Arithmetic	12.5	14.1	+1.6	3.3	3.5	+0.2	11.8	12.7	+0.9	3.4	3.4	0.0	10.0	10.8	+0.8	2.9	3.6	+0.7
Similarities	12.9	13.7	+0.8	2.0	1.8	−0.2	12.7	13.4	+0.7	2.5	2.7	+0.2	11.4	12.0	+0.6	2.4	2.1	−0.3
Vocabulary	11.8	12.3	+0.5	2.3	2.1	−0.2	11.5	12.5	+0.7	2.2	2.1	−0.1	11.4	11.7	+0.3	2.0	2.0	0.0
Picture Arrangement	10.8	13.6	+2.8	2.7	3.6	+0.9	11.6	13.8	+2.2	2.7	3.1	+0.4	10.4	11.8	+1.4	2.9	3.4	+0.5
Picture Completion	12.5	13.3	+0.8	2.1	1.9	−0.2	12.7	13.2	+0.5	2.2	2.2	0.0	10.7	11.3	+0.6	3.3	2.3	−1.1
Block Design	13.2	14.2	+1.0	3.0	2.7	−0.3	12.8	13.9	+1.1	2.9	2.6	−0.3	10.7	11.7	+1.0	2.9	2.4	−0.5
Object Assembly	11.5	13.5	+2.0	2.7	2.3	−0.4	11.9	14.0	+2.1	2.5	1.9	−0.6	10.0	12.0	+2.0	2.5	2.2	−0.3
Digit Symbol	12.4	13.9	+1.5	2.4	2.1	−0.3	12.0	13.3	+1.3	2.5	2.2	−0.3	11.7	12.3	+0.6	2.5	2.5	0.0
Scales																		
Verbal IQ	117.8	123.1	+5.3	9.1	9.4	+0.3	116.3	121.1	+4.8	11.3	10.4	−0.9	109.6	113.7	+4.1	9.6	9.7	+0.1
Performance IQ	115.7	125.9	+10.2	12.5	12.8	+0.3	115.8	126.2	+10.4	12.6	11.4	−1.2	109.0	116.8	+7.8	11.4	11.4	0.0
Full Scale	118.6	127.0	+8.5	10.4	10.6	+0.2	118.1	125.7	+7.6	11.6	10.0	−1.6	110.1	116.3	+6.2	10.1	9.9	−0.2

al background of the subjects is higher than in Wechsler's sample. It is reported and compared with Wechsler's data in Table 1. The Full Scale IQ range is normal and above normal. The mean IQ for the six-month group on the first test was 110.1. The mean IQ for the four-week group on the first test was 118.1. The mean IQ for the one-week group on the first test was 118.5. Greater variation upon retest is expected of IQ's above the mean than of normal or below normal IQ's. The results of the study, therefore, give an estimate of variation which is more than minimal.

All the six-month tests and retests were administered and scored by one of the authors and checked by another psychologist. Each four-week and one-week test and retest was administered and scored by the same examiner. These examiners were completing graduate training in psychometrics. Their training emphasized a strict methodology of administration [3] following the test manuals explicitly. The research tests were given late in the training period, after each examiner had received considerable supervised experience. Each test was checked and rescored by one of the authors.

All subjects were normal on the basis of having no clinically manifest neurosis or psychosis. None were under medical or psychological treatment at the time of testing.

RESULTS

Practice effect on the subtests can be observed in a comparison of means from test to retest. The differences between the means as well as the means and standard deviations of the weighted scores are reported in Table 2. Every mean increased significantly in the retest. The increases ranged from .3 weighted score units for Comprehension in the four-week group and Vocabulary in the six-month group, to 2.8 weighted score units for Picture Arrangement in the one-week group. There was less difference between test and retest means in the six-month group than was found in the four-week or one-week groups with the exception of Comprehension. This may reflect either the lower average IQ of the six-month group or the diminution of practice effect with a longer time interval before retest. The latter point is also suggested by the tendency of the four-week group to show less difference between means than the one-week group.

Those subtests which would seem most susceptible to simple learning showed the greatest gain in mean score. These were Digit Span, Arithmetic, Picture Arrangement, Block Design, Object Assembly, and Digit Symbol. The scores on all but Digit Span are especially susceptible to practice effect because time bonuses are allowed.

The standard deviations ranged in size from a low of 1.6 to a high of 3.6 weighted score points on the subtests. The standard deviations for the scales ranged from 9.1 to 12.6 IQ points. There were small differences in variability between the test and retest groups. Of the 33 subtest standard deviations only 4 showed a difference of more than .5. Differences ranged from 0 to 1.1. Above a .5 difference were the one-week Picture Arrangement with .9, the four-week Object Assembly with .6, and the six-month Arithmetic and Picture Completion with .7 and 1.1 respectively. The differences for the scales on the one-week and six-month groups ranged from 0 to .3. Greater differences in variability occurred in the four-week group with a range of .9 to 1.6.

Although most of these differences are small, even minor shifts become statistically significant when a high relation exists between measures as is true in a test-retest experiment. In order to test this hypothesis, the *T*-test for the significance of the difference between coefficients of relative variation [11, pp. 418-419] was applied. Since the samples were not independent, the use of the correlational term in the formula of the standard error became necessary. This decreased the size of the standard error so that even small differences between statistics became significant. When this *T*-test was applied, every subtest and scale showed a significant difference upon retest.

The Pearson product-moment correlation coefficients for each subtest and scale for each test-retest period were computed. For the subtests, weighted scores are compared. For the scales, IQ's are compared. The coefficients range downward from a high of .91 for both the Vocabulary subtest and the Full Scale IQ in the one-week retest as well as a .91 for the Full Scale IQ on the six-month retest. The lowest coefficient is .53 for Arithmetic in the four-week retest. The scales all showed test-retest coefficients above .83. The Verbal Scale was .84 at each retest interval. The Performance Scale varied from .83 to .88. The Full Scale varied from .87 to .91.

TABLE 3

THE SIGNIFICANCE OF THE DIFFERENCES* BETWEEN THE COEFFICIENTS OF CORRELATION OBTAINED WITH THREE GROUPS OF NORMAL SUBJECTS RETESTED AT ONE-WEEK, FOUR-WEEK, AND SIX-MONTH INTERVALS

Subtests	T Values†	N
Information	.58	120
Comprehension	.68	98
Digit Span	.50	98
Arithmetic	1.64	98
Similarities	1.16	120
Vocabulary	1.54	98
Picture Arrangement	.79	120
Picture Completion	.84	120
Block Design	1.50	98
Object Assembly	1.05	120
Digit Symbol	1.18	98
Scales		
Verbal IQ	0.00	120
Performance IQ	1.00	120
Full Scale IQ	1.05	120

*Determined for the differences between the two coefficients representing the extreme discrepancy among the three groups for each test.

†No T value significant below 5% level.

TABLE 4

AVERAGE TEST-RETEST COEFFICIENTS, AND STANDARD ERRORS OF MEASUREMENT FOR THE WECHSLER-BELLEVUE SUBTESTS AND SCALES (N = 158)

Subtests*	Average Correlation†	SE_m‡
Information	.86	.68
Comprehension	.74	1.21
Digit Span	.67	1.68
Arithmetic	.62	2.06
Similarities	.71	1.22
Vocabulary	.88	.73
Picture Arrangement	.64	1.82
Picture Completion	.83	.95
Block Design	.84	1.10
Object Assembly	.69	1.31
Digit Symbol	.80	1.06
*Scales**		
Verbal IQ	.84	3.96
Performance IQ	.86	4.49
Full Scale IQ	.90	3.29

*Standard errors are in terms of weighted score points for the subtests, and IQ points for the scales.

†Average of the one-week, four-week, and six-month retests, by converting obtained coefficients into Fisher's *z* functions, and reconverting the averages so determined.

‡Based on the average correlation coefficients given in the preceding column.

In order to test the significance of the differences between coefficients obtained at the several retest intervals, the *T*-test was applied. The two most disparate coefficients for each subtest and scale were converted to Fisher's z. The *T* values are reported in Table 3. None proved to be statistically significant below the 5 per cent level.

Since the time between retests was demonstrated not to affect significantly the size of the reliability coefficients, a more useful overall coefficient was determined for normal subjects by combining the results for all 158 subjects in this research. This was done by the conversion of all coefficients for each subtest and scale at each retest interval to Fisher's z, averaging z's and reconverting the averages to coefficients. The average test-retest coefficients are reported in Table 4. The most stable subtests appear to be Vocabulary with .88, Information with .86, Block Design with .84, Picture Completion with .83, and Digit Symbol with .80. The least stable subtests appear to be Object Assembly with .69, Digit Span with .67, Picture Arrangement with .64, and Arithmetic with .62. Usually a coefficient of .70 or lower is considered unsatisfactory for individual prediction. The indications are therefore that the latter subtests must be used with extreme caution when intratest clinical analysis is being undertaken on the basis of the scores on these subtests. The Full Scale is quite satisfactory with .90. The Verbal Scale with .84 and the Performance Scale with .86 are also minimally useful for individual diagnostic purposes if used with caution.

As a further measure of reliability and as a clinically useful point of reference, the standard errors of measurement were computed for each subtest and scale. The average standard errors of measurement for the combined population of 158 subjects are given in Table 4. The standard errors of the subtests are given in terms of weighted score points. They vary from

TABLE 5

THE TEST-RETEST RELIABILITY OF THE WECHSLER-BELLEVUE REPORTED IN THE LITERATURE AND IN THE PRESENT STUDY

Author	*Gibby [5]	Rabin [13]	Hamister [7]	Rabin [14]	Hamister [7]	Rabin [13]	Canter [1]	Aborn and Derner	Aborn and Derner	Canter	Aborn, Derner, and Canter
Population	Psychoneurotics	Schizophrenics	Schizophrenics	Miscellaneous Diagnoses	Miscellaneous Diagnoses	Miscellaneous Diagnoses Non-Schiz.	Multiple Sclerotics	Normals, One-Week Retest	Normals, Four-Week Retest	Normals, Six-Month Retest	Normals, Combined Average
N	32	30	34	60	53	30	47	60	60	38	158
Subtests					**Coefficients of Reliability**						
Information	.56	.89	.94	.99	.94	.89	.81	.87	.84	.87	.86
Comprehension	.20	.12	.78	.44	.76	.62	.76	.70	.74	.77	.74
Digit Span	.65	.62	.63	.77	.59	.75	.73	.67	.70	.64	.67
Arithmetic	.76	.75	.87	.68	.87	.73	.74	.57	.53	.74	.62
Similarities	.71	.38	.84	.62	.86	.79	.93	.65	.76	.70	.71
Vocabulary	.93		.90		.90		.90	.91	.88	.83	.88
Picture Arrangement	.49	.54	.78	.60	.73	.73	.86	.68	.59	.66	.64
Picture Completion	.87	.32	.68	.89	.71	.80	.73	.80	.85	.82	.83
Block Design	.87	.71	.67	.65	.71	.70	.82	.88	.84	.78	.84
Object Assembly	.53	.31	.62	.67	.66	.79	.74	.73	.62	.73	.69
Digit Symbol	.81	.34	.79	.74	.80	.91	.90	.78	.76	.85	.80
Scales											
Verbal IQ	.76	.78	.91	.89	.91	.87	.84	.84	.84	.84	.84
Performance IQ	.82	.52	.80	.76	.83	.94	.90	.88	.83	.87	.86
Full Scale IQ	.87	.55	.84	.84	.87	.89	.90	.91	.87	.91	.90

*Comparison of Forms I and II, not test-retest.

.68 for Information to 2.06 for Arithmetic. Eight of the subtests have standard errors of less than 1.50 and only three of these have standard errors of less than 1.00. The Performance Scale is the least dependable scale with a standard error of 4.49. The Full Scale is most dependable with a standard error of 3.29. The Verbal Scale has a standard error of 3.96.

The obtained coefficients for the one-week, four-week, and six-month intervals and the average for the combined group are given in Table 5. The Table also lists all other reliability coefficients on the Wechsler-Bellevue reported in the literature. There are many noticeable discrepancies between the coefficients obtained for the normal and abnormal populations. The significant differences between the results of similar abnormal populations as contrasted with the insignificant differences for the three normal populations suggests that the reliability of this test is considerably lower with abnormal subjects where the factor of the psychopathy is introduced.

SUMMARY

Various methods of determining the reliability of intelligence tests have been examined and evaluated. Inasmuch as split-half reliability is of questionable value, particularly where the splitting is done by subtests in a scale as heterogeneous as the Wechsler-Bellevue, and no truly alternate form of this test exists (although a similar test is available in Form II), the test-retest method would seem to be the most desirable reliability technique. The few Wechsler-Bellevue reliability studies available in the literature have employed abnormal subjects almost exclusively, with little attention paid to reliability obtained with normal populations.

A normal population of 158 subjects was tested and retested with the Wechsler-Bellevue Scales. The interval between retests was carefully controlled with 60 subjects retested at the end of one week, 60 subjects retested at the end of four weeks, and 38 subjects retested after a period of six months.

The mean score increased for every subtest and scale regardless of the time between retests. These increases ranged from .3 weighted score points for Comprehension in the four-week retest to 2.8 weighted score points for Picture Arrangement in the one-week retest. There seems to be a lessening of practice effect when the interval before retest is greater. Practice effect is greatest on those subtests which measure simple learning and on those in which a time factor contributes to the score.

The standard errors of measurement for the combined population of 158 subjects ranged from .68 to 2.06 weighted score points for the subtests. The standard error for the Verbal IQ is 3.96, for the Performance IQ 4.49, and 3.29 for the Full Scale IQ.

There were no significant differences between product-moment coefficients at any of the retest intervals. The lapse of time before retesting evidently does not affect the reliability, and the three groups of normal subjects were combined to yield a more useful overall coefficient for each subtest and scale. The r's ranged from .62 to .90. Arithmetic, Picture Arrangement, and Digit Span and Object Assembly are in the .60's. Similarities and Comprehension are in the .70's. Digit Symbol, Picture Completion, Block Design, Information, and Vocabulary are in the .80's. The average coefficient for the Verbal Scale is .84, for the Performance Scale .86, and for the Full Scale .90.

In the interests of maximum clinical utility, a summary table contains a complete review of all reliability studies reported in the literature on the Wechsler-Bellevue Scales along with the results of this study. It seems evident from this study and from the summarizations of other studies that the clinician must use caution in interpreting the meaning of subtest deviations. In many cases, the degree of reliability is too low to warrant more than the most tenuous hypotheses. The scales, however, and particularly the Full Scale, can be considered diagnostically dependable with normal subjects.

Received February 6, 1950
Early publication.

REFERENCES

1. CANTER, A. H. Direct and indirect measures of psychological deficit in multiple sclerosis. Unpublished Ph.D. dissertation, Columbia Univ., 1949.
2. CUMMINGS, S. B., JR., MACPHEE, H. M., AND WRIGHT, H. F. A rapid method of estimating the IQ's of sub-normal white adults. *J. Psychol.*, 1946, **21**, 81-90.

3. DERNER, G. F. AND ABORN, M. *The clinical diagnosis of intelligence.* Mimeographed. New York: Authors, 1949.
4. GEIL, G. A. A clinically useful abbreviated Wechsler-Bellevue Scale. *J. Psychol.*, 1945, **20**, 101-108.
5. GIBBY, R. G. A preliminary survey of certain aspects of Form II of the Wechsler-Bellevue Scale as compared to Form I. *J. clin. Psychol.*, 1949, **5**, 165-169.
6. GURVITZ, M. S. An alternate short form of the Wechsler-Bellevue test. *Amer. J. Orthopsychiat.*, 1945, **15**, 727-733.
7. HAMISTER, R. C. Test-retest reliability of the Wechsler-Bellevue. *J. consult. Psychol.*, 1949, **13**, 39-44.
8. HUNT, W. A., KLEBANOFF, S. G., MENSH, J. A., AND WILLIAMS, M. The validity of some abbreviated individual intelligence scales. *J. consult. Psychol.*, 1948, **12**, 48-52.
9. PATTERSON, C. H. A comparison of various "short forms" of the Wechsler-Bellevue Scale. *J. consult. Psychol.*, 1946, **10**, 260-267.
10. PATTERSON, C. H. A further study of two short forms of the Wechsler-Bellevue scale. *J. consult. Psychol.*, 1948, **12**, 147-152.
11. PEATMAN, J. G. *Descriptive and sampling statistics.* New York: Harper, 1947.
12. RABIN, A. I. A short form of the Wechsler-Bellevue test. *J. appl. Psychol.*, 1943, **27**, 320-324.
13. RABIN, A. I. Fluctuations in the mental level of schizophrenic patients. *Psychiat. Quart.*, 1944, **18**, 78-91.
14. RABIN, A. I. Test constancy and variation in the mentally ill. *J. gen. Psychol.*, 1944, **31**, 231-239.
15. SPRINGER, N. N. A short form of the Wechsler-Bellevue intelligence test as applied to Naval personnel. *Amer. J. Orthopsychiat.*, 1946, **16**, 341-349.
16. WECHSLER, D. *The measurement of adult intelligence.* (3rd Ed.) Baltimore: Williams & Wilkins, 1944.
17. WECHSLER, D. *The Wechsler-Bellevue intelligence scale. Form II. Manual for administering and scoring the test.* New York: Psychological Corp., 1946.

THE DIAGNOSTIC SIGNIFICANCE OF AN OBJECTIVE MEASURE OF INTRATEST SCATTER ON THE WECHSLER-BELLEVUE INTELLIGENCE SCALE

JULES D. HOLZBERG AND MAURICE A. DEANE
CONNECTICUT STATE HOSPITAL

SCATTER, or the manner in which a testee distributes successes and failures on an intelligence test, has long been used by psychologists in diagnostic studies. Although scatter had been used very early to describe aberrations in intellectual functioning on such tests as the Stanford-Binet, it received its greatest impetus as a diagnostic technique from the publication of the Wechsler-Bellevue Intelligence Scale [7] and especially from the work of Rapaport [6]. However, most attention to scatter on the Wechsler-Bellevue has been directed to the problem of scatter between subtests of this scale, although clinicians have utilized scatter within a subtest qualitatively to arrive at diagnostic conclusions [1, p. 29]. The purpose of the present study was to determine whether scatter within subtests of the Wechsler-Bellevue can be quantitatively expressed and whether such a measure validly contributes to the diagnostic understanding of psychiatric patients.

For the purpose of this study, *"intratest scatter"* is used to designate scatter within a subtest as opposed to "intertest scatter" which is now in common clinical usage in the analysis of relationship between subtests.

Several reviews of the literature on scatter have been published. The present paper will not attempt any such review, but will refer only to work of specific relationship to this study. In their review of quantitative measures of scatter on the Stanford-Binet, Harris and Shakow [2, 3] found no clear evidence that such quantitative measures were valid, although they felt that quantitative scatter beyond certain limits might be of pathological significance. Hunt and Cofer [4] are in essential agreement that scatter on the Stanford-Binet appeared to have little clinical significance. Nevertheless, as Mayman [5] indicates, there still remained the clinical impression that scatter on the Stanford-Binet was of diagnostic importance. With publication of the Wechsler-Bellevue Intelligence Scale, the use of scatter as a diagnostic technique advanced and there evolved clinical and statistical evidence that intertest scatter on this scale could discriminate between various psychiatric groups [6, 7].

Intratest scatter, however, has received little research attention although it is used frequently in current clinical practice. Although scatter on the Stanford-Binet resembles intratest scatter as defined in this study, it differs in that it does not measure scatter within a specific psychological function.

Rapaport has attempted the use of intratest scatter on the Wechsler-Bellevue in terms of its diagnostic significance. In his work, he classifies the items of several subtests into easy and hard items and attempts to relate failure on easy items with concurrent success on hard items to various psychiatric disorders. On several subtests, he rearranges the items to conform to an order of difficulty suggested by his data, and as a result, offers an order of difficulty differing from that of Wechsler. Such a rearrangement, while unquestionably useful if valid, introduces difficulty in the clinician's analysis of intratest scatter if for no other reason than the practical one that the test blanks which he uses to examine scatter conform to Wechsler's order. In the present study, no attempt was made to rearrange the items. The arrangement of items as currently used by clinicians served as the basis for determining whether intratest scatter could be measured and whether such measures discriminated between various psychiatric groups.

PROCEDURE

In order to devise a satisfactory measure of intratest scatter, a workable definition of intratest scatter had to be established. To attain this, six members of the psychology staff of the Connecticut State Hospital were given four lists of scores on a hypothetical test whose items were arranged in increasing order of difficulty, and they were asked to arrange the four lists in order of the extent of intratest scatter. They were then questioned as to the rationale for the order they had given. The four scatters are listed below:

Item	*Scatter 1*	*Scatter 2*	*Scatter 3*	*Scatter 4*
1	+	+	−	−
2	−	−	−	−
3	−	+	−	−
4	−	−	−	−
5	−	+	−	−
6	−	−	−	+
7	−	+	+	−
8	−	−	−	−
9	+	+	+	+

Considerable disagreement was found among the six raters and subsequent discussion with other clinicians has confirmed that psychologists differ in their evaluation of intratest scatter. Analysis of the reasons given for the orders of arrangement revealed three distinct criteria for measuring intratest scatter. The first was based on the number of incorrect items before the last correct item is reached. Three of the raters followed this rationale and felt that Scatter 1 showed more scatter than Scatter 2. They reasoned that one may consider scatter in terms of probability, and that one would much more expect Scatter 2 to occur than Scatter 1.

The second criterion was based on the concept of number of changes, i.e., the more changes from correct to incorrect, and vice versa, the greater the amount of scatter. Therefore, using this standard, the other three raters felt that Scatter 2 showed more scatter than Scatter 1.

The third criterion was concerned with the location of change. Considering Scatter 3 and Scatter 4 where the number right, the number of changes, and the last items attempted are the same in both cases, the three raters who thought Scatter 1 scattered more than Scatter 2 extended their rule of probability and felt that Scatter 3 scattered more than Scatter 4. The other three maintained that Scatter 4 showed the greater scatter because there was a greater distance between the two changes in Scatter 4, that is, the change was "maintained" for a longer interval.

The first criterion was selected because it seemed the simplest to express quantitatively. In constructing the formula, the assumption was made that the subtest items of the Wechsler-Bellevue Scale were arranged in order of increasing difficulty. That this assumption might not be valid did not seem of practical importance, for the scatter caused by any items not in absolute order of difficulty would be common to all testees and thus be a constant factor.

After discarding many proposed formulae, one was found which combined as many of the properties of scatter (using the first criterion) as was possible. That formula was

$$S = \frac{N - N'}{T}$$

where

S = the scatter coefficient
N = the last item for which any credit was given
N′ = the total number of items for which *any* credit was given
T = total raw score

The numerator of this formula can be defined as the number of totally incorrect items within the range of items receiving any credit. The denominator served to differentiate scatters in those subtests where partial credit was given. For example, in the following two scatters, both would have the same value for the numerator ($N - N' = 4 - 3 = 1$), but Scatter 6, having a smaller value for T, would have the larger scatter coefficient and thus would indicate greater scatter.

Item	*Scatter 5*	*Scatter 6*
1	3	3
2	3	1
3	0	0
4	3	3
	T = 9	T = 7

$$\text{Scatter 5} = \frac{N - N'}{T} = \frac{4 - 3}{9} = \frac{1}{9} = .11$$

$$\text{Scatter 6} = \frac{N - N'}{T} = \frac{4 - 3}{7} = \frac{1}{7} = .14$$

Fifteen scatter measures were utilized for the subtests of the Wechsler-Bellevue Scale. They were:

1. Intratest scatter coefficient for *Information* (test of 25 items).

2. Intratest scatter coefficient for *Comprehension* (test of 10 items).

3. Intratest scatter coefficient for *Similarities* (test of 12 items).

4. Intratest scatter coefficient for accuracy on *Arithmetic*. Here time credit was eliminated, and the test consisted of ten items. For example:

Item	*Time Credit*	*Accuracy Credit*	*Total Credit*
1	0	1	1
2	0	1	1
3	0	1	1
4	0	1	1
5	0	1	1
6	0	1	1
7	0	1	1
8	0	0	0
9	1	1	2
10	0	1	1
	—	—	—
	1	T = 9	10

$$S = \frac{N - N'}{T} = \frac{10 - 9}{9} = \frac{1}{9} = .11$$

The total raw score (T) for arithmetic is the accuracy credit.

5. Intratest scatter coefficient for *Digit Span Forward*. This test consisted of 14 items, there being two items for each of the seven digit span levels, and both trials on a digit span level were assumed correct if the first of each series was answered correctly. For example:

Item	*Series*	*Score*
1	582	+
2	694	+ (not given)
3	6439	—
4	7286	+
5	42731	—
6	75836	—

$$S = \frac{N - N'}{T} = \frac{4 - 3}{4} = \frac{1}{4} = .25$$

The raw score (T) earned for the above is 4 (the highest level of success is four digits).

6. Intratest scatter coefficients for *Digit Span Backward*. (Test of 12 items handled the same as Digit Span Forward).

7. Intratest scatter coefficient for *Vocabulary* (test of 42 items).

8. Intratest scatter coefficient for *Picture Completion* (test of 15 items).

9. Intratest scatter coefficient for accuracy on *Picture Arrangement*. Here the test consisted of 15 items which is the maximum credit for accuracy on all of the six sets of pictures with time credit eliminated. For example:

Items	Cards	Time Credit	Accuracy Credit	Total Credit
1, 2	House	0	2	2
3, 4	Hold Up	0	2	2
5, 6	Elevator	0	0	0
7, 8, 9	Flirt	0	2	2
10,11,12	Taxi	1	2	3
13,14,15	Fish	0	0	0
		—	—	—
		1	T = 8	9

N = 11; the last item correct was #11 or the second of three credits given for accuracy on "Taxi."

N′ = 8; items (1, 2, 3, 4, 7, 8, 10, and 11) were answered correctly.

T = 8; the number of credits received for accuracy alone. (For this scale, N′ will always have the same value as T).

$$S = \frac{N - N'}{T} = \frac{11 - 8}{8} = \frac{3}{8} = .38$$

10. Intratest scatter coefficients for accuracy on *Object Assembly*. Here the tests consists of 18 items, the maximum number of credits for accuracy with time credit eliminated. For example:

Items	Objects	Time Credit	Accuracy Credit	Total Credit
1,2,3,4,5,6	Manikin	0	6	6
7,8,9,10,11,12	Profile	1	5	6
13,14,15,16,17,18	Hand	2	6	8
		—	—	—
		3	T = 17	20

N = 18; the last item correct was #18, or the sixth of six credits for "Hand."

N′ = 17; item #12 was answered incorrectly.

T = 17; the number of credits received for accuracy alone. (As in Picture Arrangement, the value of N′ will always be the same as the value of T).

$$S = \frac{N - N'}{T} = \frac{18 - 17}{17} = \frac{1}{17} = .06$$

11. Intratest scatter coefficient for *Block Design-Accuracy*. Here the test consisted of 21 items, with three items for each card and time credit eliminated. For example:

Items	Card	Time Credit	Accuracy Credit	Total Credit
1,2,3	1	2	3	5
4,5,6	2	1	3	4
7,8,9	3	1	3	4
10,11,12	4	1	3	4
13,14,15	5	0	0	0
16,17,18	6	1	3	4
19,20,21	7	0	3	3
		—	—	—
		6	18	24

N = 21; the last item correct was #21, or the third of three credits given for card 7.

N′ = 18; items #13, #14, and #15 (card 5) were incorrect. (Here again, as in Picture Arrangement, the value of N′ is always the same as the value for T).

T = 18; the number of credits for accuracy alone.

$$S = \frac{N - N'}{T} = \frac{21 - 18}{18} = \frac{3}{18} = .17$$

12. Intratest scatter coefficient for *Block Design-Time*. Here the test consists of 21 items, the maximum number of time credits possible, with credit for accuracy eliminated. This coefficient was computed in the same way as that for Block Design-Accuracy. For example, using the same data as for Block Design-Accuracy above:

N = 16; the last item correct was #16, or the first of three time credits given on card 6.

N′ = 6; items #1, #2, #4, #7, #10, and #16 were correct.

T = 6; the number of credits for time. (The value of N′ will always be the same as that for T).

$$S = \frac{N - N'}{T} = \frac{16 - 6}{6} = \frac{10}{6} = 1.67$$

13. A second intratest scatter coefficient for *Block Design-Time* (test of seven items). For example:

Item	Time Credit	Accuracy Credit	Total Credit
1	2	3	5
2	1	3	4
3	1	3	4
4	1	3	4
5	0	0	0
6	1	3	4
7	0	3	3
	T = 6	18	24

N = 6; the last item for which any time credit was given was #6.

N′ = 5; items #1, #2, #3, #4, and #6 were all given at least partial credit.

T = 6; the number of credits for time.

$$S = \frac{N - N'}{T} = \frac{6 - 5}{6} = \frac{1}{6} = .17$$

14. Intratest scatter coefficient for *Digit Symbol* (test of 67 items). Here N′ was equal to the number of items for which any credit was given (one-half credits were considered correct).

15. A second intratest scatter coefficient for *Digit Symbol* (list of 67 items). Here N′ was equal to the number of items for which full credit was given (one-half credits were considered incorrect).

SUBJECTS

Three psychiatric groups, i.e., neurotic, organic, and schizophrenic, were chosen as subjects for the study. Five limiting factors were placed on the original sample selected:

1. Age: only individuals below the age of 50 were selected in order to minimize the effects of senescence on the test results.
2. Intelligence: all those whose optimal intelligence (premorbid level of intellectual functioning) was dull normal or higher were included, provided a present functioning intelligence quotient of at least 65 was demonstrated.
3. Weighted Scores: only those records where the weighted score of each subtest was equal to at least one were used.
4. Education: six years of elementary school was the minimum of education for any member of the sample.
5. Representativeness: only those records which were considered representative, i.e., in which attention, concentration, and effort were satisfactory, were used.

Twelve neurotics and eight organics met all these criteria. Records of forty schizophrenic patients who met all of the five criteria were available, and it was thus possible, in the comparison of the schizophrenics with each of the other two groups, to match members of the sample for age, optimal intelligence, and approximately for education. In comparing the organic group with the neurotic group, matching was impossible because of the limited number of cases. The neurotic group was therefore reduced to eight individuals, the four records eliminated serving to approximately equate the two groups for age and optimal intelligence.

STATISTICAL PROCEDURE

In comparing one group with another, means for the fifteen different intratest scatters were found for each group. The difference between the means for each intratest scatter was tested for significance. Because of the small number of subjects, Fisher's small sampling technique was used.

FINDINGS

Schizophrenics and Neurotics. Twelve schizophrenics matched with twelve neurotics for age, optimal intelligence, and education were compared for intratest scatter. The mean age for the schizophrenic group was 29.0 with a range of 21 to 47 and the neurotic mean age was 29.5 with a range of 19 to 44. The mean educational grade completed for both groups was 10.5. The range of completed education was 6th grade to completion of college for the

neurotics and 6th grade to completion of two years' postgraduate study for the schizophrenics. Optimal intellectual functioning for both groups ranged from dull normal to very superior with a median of high average. The diagnostic composition of the two groups is given in Table 1.

TABLE 1
DIAGNOSTIC COMPOSITION OF SCHIZOPHRENICS AND NEUROTICS

Schizophrenics	No.	Neurotics	No.
Hebephrenic	4	Reactive Depression	4
Simple	3	Anxiety	3
Paranoid	2	Hysteria	2
Catatonic	2	Psychasthenia	1
Other Types	1	Neurasthenia	1
		Mixed Type	1
Total	12	Total	12

Table 2 gives the means and standard deviations of each of the fifteen intratest scatters for schizophrenics and neurotics, and Fisher's *t*'s obtained after comparing these two clinical groups. Using a *t* of 2.074 as being significant at the five per cent level (Fisher's tables), it is apparent that Comprehension, Picture Completion, and Block Design-Accuracy scatters are significantly greater for schizophrenics than for neurotics. All other differences in intratest scatters on other subtests are not significant.

TABLE 2
SIGNIFICANCE OF DIFFERENCES IN INTRATEST SCATTER BETWEEN SCHIZOPHRENICS AND NEUROTICS

	Schizophrenics		Neurotics		
	M	SD	M	SD	*t*
Information	.40	.33	.41	.17	.09
Comprehension	.58	.59	.16	.19	2.22
Similarities	.27	.33	.20	.17	.63
Arithmetic	.19	.28	.16	.22	.28
Digit Span Forward	.09	.14	.07	.13	.28
Digit Span Backward	.10	.12	.08	.10	.34
Vocabulary	.25	.11	.23	.16	.35
Picture Completion	.59	.42	.27	.26	2.11
Picture Arrangement	.22	.39	.15	.14	.55
Object Assembly	.08	.13	.02	.04	1.43
Block Design-Accuracy	.20	.29	.01	.05	2.11
Block Design-Time-1	1.29	1.28	.88	.77	.90
Block Design-Time-2	.25	.36	.10	.17	1.27
Digit Symbol-1	.00	.00	.00	.01	1.00
Digit Symbol-2	.04	.08	.03	.06	.37

Schizophrenics and Organics. Eight schizophrenics were matched with an equal number of organics for age, education, and optimal intelligence. The mean age for the schizophrenic group was 37.5 as compared to 37.6 for the organics with a range of 28 to 49 for both groups. The range in educational accomplishment was 6th grade to completion of 3rd year high school for both groups, with the mean for the organics at 8.4 and for the schizophrenics at 8.3. The optimal levels of intelligence ranged from dull normal to bright normal with a median of average for both groups. Table 3 lists the diagnostic classifications of both groups.

TABLE 3
DIAGNOSTIC COMPOSITION OF SCHIZOPHRENICS AND ORGANICS

Schizophrenics	No.	Organics	No.
Simple	4	C.N.S. Syphilis	7
Catatonic	1	Psychosis due	
Hebephrenic	1	to trauma	1
Paranoid	1		
Other Types	1		
Total	8	Total	8

The means, standard deviations, and Fisher's *t*'s for this phase of the study are reported in Table 4. Using Fisher's tables, a *t* of 2.145 was considered significant at the five per cent level, and here Information, Comprehension, and Block Design-Accuracy show significantly greater scatter for schizophrenics. All other differences are not significant.

TABLE 4
SIGNIFICANCE OF DIFFERENCES IN INTRATEST SCATTER BETWEEN SCHIZOPHRENICS AND ORGANICS

	Schizophrenics		Organics		
	M	SD	M	SD	*t*
Information	.46	.30	.21	.13	2.18
Comprehension	.50	.48	.13	.09	2.16
Similarities	.85	.95	.22	.37	1.62
Arithmetic	.12	.14	.11	.17	.12
Digit Span Forward	.11	.13	.05	.08	1.12
Digit Span Backward	.09	.11	.07	.13	.23
Vocabulary	.19	.15	.32	.17	1.51
Picture Completion	.70	.44	.41	.35	1.38
Picture Arrangement	.39	.49	.35	.64	.13
Object Assembly	.15	.24	.03	.05	1.28
Block Design-Accuracy	.36	.33	.00	.00	2.77
Block Design-Time-1	1.05	1.91	.46	.61	.79
Block Design-Time-2	.27	.66	.07	.10	.78
Digit Symbol-1	.00	.00	.06	.08	1.87
Digit Symbol-2	.05	.07	.16	.13	1.93

Organics and Neurotics. Eight organics and eight neurotics were approximately equated as groups. The mean age for the organics was

37.6 with an age range from 28 to 49. The mean age for the neurotic group was 32.4, the range being from 25 to 44. The educational level for the organics ranged from 6th grade to 3rd year high school with a mean of 8.4. The neurotics ranged from 6th grade to first year college with a mean of 10.0. For both groups, the median optimal intelligence level was average with a range from dull normal to bright normal. The diagnostic breakdown for both groups is given in Table 5.

TABLE 5
DIAGNOSTIC COMPOSITION OF ORGANICS AND NEUROTICS

Organics	No.	Neurotics	No.
C.N.S. Syphilis	7	Reactive Depression	4
Psychosis with Trauma	1	Anxiety	2
		Hysteria	1
		Neurasthenia	1
Total	8	Total	8

Utilizing Fisher's tables, a t of 2.145 was considered significant at the five percent level. Differences in intratest scatter were found significant for Information and Digit Symbol-2 (Table 6). With Information, scatter was greater for neurotics, whereas with Digit Symbol-2, scatter was greater for organics. There were no other significant differences.

TABLE 6
SIGNIFICANCE OF DIFFERENCES IN INTRATEST SCATTER BETWEEN ORGANICS AND NEUROTICS

	Organics		Neurotics		
	M	SD	M	SD	t
Information	.21	.13	.46	.16	3.21
Comprehension	.13	.09	.22	.20	1.08
Similarities	.22	.37	.25	.18	.19
Arithmetic	.11	.17	.22	.24	1.00
Digit Span Forward	.05	.08	.09	.14	.77
Digit Span Backward	.08	.13	.10	.11	.42
Vocabulary	.32	.17	.25	.16	.80
Picture Completion	.41	.35	.31	.30	.59
Picture Arrangement	.35	.64	.19	.16	.64
Object Assembly	.03	.05	.03	.02	.00
Block Design-Accuracy	.00	.00	.00	.00	.00
Block Design-Time-1	.46	.61	.85	.90	.95
Block Design-Time-2	.07	.10	.10	.20	.31
Digit Symbol-1	.06	.08	.00	.01	1.74
Digit Symbol-2	.16	.13	.03	.03	2.70

Composite Findings. Considering the eleven scales of the Wechsler-Bellevue, the following findings are noted:

1. *Information*: Both the neurotic and schizophrenic groups scattered significantly more than the organic group, but differences between neurotics and schizophrenics were not significant.

2. *Comprehension*: The schizophrenics were found to have a significantly greater intratest scatter than neurotics and organics, whereas differences between the two latter groups were not reliable.

3. *Digit Span*: No significant differences between the three groups were found.

4. *Arithmetic*: There were no significant differences between the three groups.

5. *Similarities*: No significant differences between the three groups were found.

6. *Vocabulary*: There were no significant differences between the three groups.

7. *Picture Arrangement*: No significant differences between the three groups were found.

8. *Picture Completion*: The schizophrenic group scattered significantly more than the neurotic group but not more than the organic group. No real difference between neurotics and organics was found.

9. *Block Design*: There was a significantly greater scatter shown on Block Design-Accuracy by the schizophrenic group as compared to the neurotic and organic groups but there were no differences between the latter two groups. Block Design-Time failed to show any real differences between the three groups.

10. *Object Assembly*: No significant differences between the three groups were found.

11. *Digit Symbol*: The organic group showed significantly greater scatter than the neurotic group when test items which were given half credit were considered incorrect. No other significant differences were found.

DISCUSSION

Wechsler in his analysis of various psychopathological groups found that intertest scatter is small for organics, moderate for neurotics, and marked for schizophrenics [7]. It has generally been assumed in clinical practice that *intratest* scatter paralleled *intertest* scatter as reported by Wechsler. The schizophrenic is considered the most unpredictable testee in terms of his pattern of successes and failures. The neurotic reveals fluctuating performance as a consequence of anxiety, although not displaying intratest scatter to the same extent as the schizophrenic. The organic, with his intellectu-

al and associational poverty, shows lowered performance on intelligence tests, with greater consistency in intratest patterning. The findings of the present study on intratest scatter tend to confirm these clinical impressions particularly in regard to schizophrenics (Table 7). However, intratest scatter as measured in this study shows little significant difference between neurotics and organics, except that the neurotics reveal greater scatter on the information subtest and the organics demonstrate greater scatter on digit symbol.

TABLE 7

STATISTICALLY SIGNIFICANT INTRATEST SCATTER BETWEEN SCHIZOPHRENICS, NEUROTICS, AND ORGANICS

Group Comparisons			Subtests
Schizophrenic	>	Neurotic	Comprehension, Picture Completion, Block Design
Schizophrenic	>	Organic	Information, Comprehension, Block Design
Neurotic	>	Organic	Information
Organic	>	Neurotic	Digit Symbol

On six Wechsler-Bellevue subtests, no significant differences in intratest scatter were found between any of the three groups. This would suggest that intratest scatter resulting from psychopathology may not be significant in all areas of intellectual functioning measured by the Wechsler-Bellevue. On the other five subtests (information, comprehension, picture completion, block design, and digit symbol), significant differences were found. In no instance did neurotics and organics show significantly greater scatter than schizophrenics. The only discrepant finding in terms of the expected order of extent of intratest scatter is on the digit symbol test where the organic group showed greater intratest scatter than the neurotics. This discrepant finding, however, is an artifact produced by the method of scoring that was adopted for this study. It was found that when one-half credits on this subtest were considered correct, no significant difference in scatter was demonstrated between neurotics and organics. When, however, these one-half credits were considered incorrect responses, a significant difference was found. This tends to confirm the clinical finding that organics frequently perseverate with the half credit response.

A comparison of the present findings with those of Rapaport are of interest since his is the only known attempt to apply diagnostically intratest scatter on the Wechsler-Bellevue, although this is done on an essentially qualitative level. Rapaport uses no organic population, so no comparison with that group is possible. However, he does use schizophrenics and neurotics although he has a much finer breakdown of psychiatric classifications than was possible in this study.

Table 8 lists the seven subtests for which Rapaport discusses intratest scatter, which is defined by him in terms of failing easy items and passing hard items. On vocabulary, the present study finds no significant difference in intratest scatter between schizophrenics and neurotics, whereas Rapaport finds this scatter characteristic of schizophrenics and certain neurotics. Actually, the results are not very contradictory because Rapaport is merely stating that intratest scatter of schizophrenics and certain neurotics is greater than that of his normals, a fact that the present study is unable to verify because of the absence of a comparative normal population. Of interest, however, is the fact that his schizophrenic population and some neurotics show little difference between each other, and the present study finds no difference in intratest scatter between the schizophrenics and the total neurotic group. Comprehension shows clear agreement between the results of Rapaport and those of the present study. With information, a similar situation exists as with vocabulary. Rapaport finds greater intratest scatter for schizophrenics and some neurotic groups as compared to normals and the findings of the present study confirm the fact that there is an absence of significant difference between schizophrenics and neurotics. On similarities, a marked difference in findings is shown. Rapaport finds that schizophrenic and neurotic depressives show greater scatter than other neurotics whereas the results of the present study do not demonstrate any significant difference between these two groups. In arithmetic, the results of both studies are again in disagreement. Rapaport finds a difference between schizophrenics and neurotics, a difference which is not revealed in the present study. On the picture arrangement test, Rapaport again finds schizophrenics and some neurotics showing greater intratest scatter than normals. The results of

the present study merely indicate that no statistically significant difference in intratest scatter was obtained between schizophrenics and neurotics. On the picture completion test, findings of the present study are in complete agreement with that of Rapaport.

TABLE 8
RAPAPORT'S FINDINGS ON INTRATEST SCATTER WITH SCHIZOPHRENICS AND NEUROTICS COMPARED WITH FINDINGS OF PRESENT STUDY

Subtest*	Present Findings	Rapaport's Findings
Information	No significant difference	Characteristic of schizophrenics and certain neurotics.
Comprehension	Schizophrenics > Neurotics	Characteristic of schizophrenics.
Similarities	No significant difference	Characteristic of schizophrenics and severe neurotic depressives; not characteristic of other neurotics.
Arithmetic	No significant difference	Characteristic of schizophrenics; not characteristic of neurotics.
Vocabulary	No significant difference	Characteristic of schizophrenics and certain neurotics.
Picture Completion	Schizophrenics > Neurotics	Characteristic of schizophrenics; not characteristic of neurotics.
Picture Arrangement	No significant difference	Characteristic of schisophrenics and depressive neurotics.

*These seven subtests are the only ones for which Rapaport presents findings on intratest scatter.

Of the seven subtests wherein it has been possible to compare Rapaport's study with the present study, comprehension and picture completion reveal findings in complete agreement with those of Rapaport, and partial agreement is found with vocabulary, information, and picture arrangement. The discrepancies that do appear in the findings may well be the result of the fact that Rapaport has rearranged the items of subtests in many instances to conform to his found order of difficulty, has used a finer and more homogeneous classification of diagnostic groups than was possible here, and has used a concept of scatter which is not identical with that developed in this study.

The application of the measure devised for intratest scatter to more homogeneous psychopathological groups may well sharpen the differences that were not apparent in the present study. No attempt was made to rearrange the items of the subtests to conform to a presumed order of difficulty since the arrangement of items offered by Wechsler is that which is used in current clinical practice. To accept Rapaport's findings of order of difficulty would imply acceptance of his control population as "normal," an assumption the writers were not prepared to make. To change Wechsler's order of difficulty would alter the concept of intratest scatter as it is usually used by clinicians. It was felt that whatever discrepancies that were present in Wechsler's order of difficulty would affect all groups equally and this factor would then be constant. However, with a more accurate scaling of items, the present technique devised for measuring intratest scatter might more sharply discriminate between groups, and perhaps show significant differences where none was demonstrated.

It should be emphasized that the absence of significant differences in intratest scatter on certain subtests between the three pathological groups used does not imply that these three groups do not scatter more than normals. Since a comparable normal population was not available, it is not possible to designate what intratest scatter for normals would be. It is therefore feasible that comparison of these three groups with a normal population might demonstrate that, although the three groups do not show significant differences in intratest scatter among themselves on certain subtests, these three groups show significantly greater intratest scatter than normals on these same subtests.

Certain precautions in the interpretation of the results should be emphasized. Clearly, the size of the groups used were too small to permit indiscriminate application of these findings to clinical cases. The age range was rather large for the sample used, and it would have been desirable to limit the subjects used to more homogeneous diagnostic groups. Control data on the extent of intratest scatter within a normal population are highly desirable and essential. In spite of these limiting factors, however, the present study does offer a new approach to the measurement of intratest scatter and is deserving of further study.

It is not the authors' purpose to offer an objective measure of intratest scatter that will

ever eliminate the need for disciplined clinical judgment in evaluating intratest scatter. The measure that has been devised is merely conceived as an adjunct to present clinical practice and to the extent that this measure helps to sharpen clinical practice, to that extent will it serve as a contribution to the field of diagnostic testing. It may also be useful in research where there is need to measure the extent of, or change in, intratest scatter.

SUMMARY

An objective measure of intratest scatter on the Wechsler-Bellevue subtests was devised and applied to a group of neurotics, organics, and schizophrenics. Intratest scatter of schizophrenics was found to be significantly greater than that of neurotics on comprehension, picture completion, and block design. Schizophrenic scatter was also reliably greater than organic scatter on information, comprehension, and block design. When neurotics were compared with organics, neurotics scattered significantly more than organics on information, and less than organics on digit symbol when one-half credits on this test were considered incorrect.

Received July 20, 1949.

REFERENCES

1. BURTON, A. AND HARRIS, R. E. *Case histories in clinical and abnormal psychology.* New York: Harper, 1947.
2. HARRIS, A. J. AND SHAKOW, D. The clinical significance of numerical measures of scatter on the Stanford-Binet. *Psychol. Bull.*, 1937, **34**, 134-150.
3. HARRIS, A. J. AND SHAKOW, D. Scatter on the Stanford-Binet in schizophrenics, normal, and delinquent adults. *J. abnorm. soc. Psychol.*, 1938, **33**, 100-111.
4. HUNT, J. MCV. AND COFER, C. N. Psychological deficit. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* New York: Ronald, 1944, pp. 971-989.
5. MAYMAN, M. Review of the literature on "scatter." In D. Rapaport, M. Gill, and R. Schafer, *Diagnostic psychological testing.* Chicago: Year Book Publishers, 1945, pp. 548-558.
6. RAPAPORT, D., GILL, M. AND SCHAFER, R. *Diagnostic psychological testing.* Chicago: Year Book Publishers, 1945.
7. WECHSLER, D. *The measurement of adult intelligence.* (3rd Ed.). Baltimore: Williams & Wilkins, 1944.

PERSONALITY PROBLEMS OF THE SPINAL CORD INJURED[1]

ALFRED D. MUELLER
VETERANS ADMINISTRATION HOSPITAL, MEMPHIS, TENN.

A NUMBER of studies have been made in recent years concerning personality characteristics, emotional conflicts, feelings and attitudes, and general adjustment problems of the spinal cord injured. Three such studies made during the past year are selected as the basis for appraising personality characteristics of paraplegic patients and for evaluating the psychological factors involved in the rehabilitation process. They were selected mainly because they utilize different measurement techniques in studying these patients.

Mosak [2] studied a group of twenty-five paraplegic patients in a Midwest VA Paraplegic Center, and an equal number of non-paraplegic veterans as controls. He compared the two groups on the basis of test results of the Harrower-Erickson Group Rorschach and the Draw-a-Person test.

McLaughlin [1] studied a group of twenty-five paraplegic patients in two Eastern VA Paraplegic Centers. He based his study on the test results of the Wechsler-Bellevue Scale and the Thematic Apperception Test.

Mueller and Thompson [3] studied an unselected group of 140 paraplegic patients in a Southern VA Paraplegic Center. They employed tests of mental ability and of occupational interests as means of stimulating, directing, and motivating rehabilitation activities.

Using the method of differences of means, Mosak found certain statistically significant differences in personality structure between his two groups. First, he found that the paraplegic group gave a larger number of bad responses on the Rorschach test. The *M*, *CF*, and *F*+ differences were significant at the 1 per cent level; *C*, *Y*, and *F*— differences were significant at the 2 per cent level; and *Sum-C* difference was significant at the .1 per cent level. Third, differences in content were found as follows: Blood, Fire, Anatomy, and Card Description were significant at the 5 per cent level; and Dirt and Human responses were significant at the 1 per cent level.

According to Beck, human movement responses (*M*) are representative of very deep wishes, innermost psychic activity, kinaesthetically determined. Hence, it is hypothesized that the reduction in human movement responses in the paraplegic group is due to reduction in kinaesthetic stimulation. The Draw-a-Person test gave some support to this hypothesis, since only six of the paraplegic group drew persons whose legs were functional, while twelve of the control group drew figures in which movement seemed apparent.

The reduction in human responses (*H*) shows that paraplegics tend to lack relationship with other people. There is poor human identification or empathy in a paraplegic group and less facility for forming associations with other people. This characteristic is also shown in the affective sphere. The paraplegic responds primarily to emotional elements of his environment and is easily motivated by external stimuli. His most prevalent emotional pattern is one of adolescent egocentricity, and there are many elements of infantile, uncontrolled affective behavior. Manifestations of this uncontrolled emotionality and potential explosiveness appear in the number of pure Color responses and of Fire responses.

More shading, or *Y* responses, were elicited from the paraplegic than from the control group. The difference is significant at the 2 per cent level. This indicates a tendency to anergic states and passivity as well as dys-

[1]From Psychology Section, Veterans Administration Medical Teaching Group, Kennedy Hospital, Memphis, Tennessee. Sponsored by the Veterans Administration and published with permission of Chief Medical Director, Department of Medicine and Surgery, Veterans Administration, who assumes no responsibility for the opinions expressed or the conclusions drawn by the author.

phoric or depressive moods.

Specific personal needs are reflected in content materials. The paraplegic records show many Anatomy, Blood, and Dirt responses. These "reflect their constant preoccupation with their injuries, with the incessant treatment, and with the 'messiness' of their conditions, especially with respect to their incontinence."

Mosak states that "the paraplegic may be described as one who is motivated with greater ease by stimuli external to himself and whose emotional makeup is primarily egocentric. However, the potentiality for mature emotional response as well as for violent explosiveness is still present. He is intensely preoccupied with his own condition and the 'messiness' of his life, is aware of the inutility of his legs constantly, but knowing that protest will not alter his condition, has become passive in his relationships. He is in many instances depressed and listless. The facility for association with other people has decreased, but he still observes group mores."

McLaughlin undertook to investigate emotional conflicts, adjustment difficulties, and personality characteristics and attitudes of paraplegic patients as revealed by twenty TAT pictures. The themas were analyzed for conscious content on seven categories considered applicable to paraplegics, as follows: (1) attitude toward health and strength; (2) social adjustments; (3) emotional maturity; (4) emotional tone; (5) role in life; (6) attitude toward sexual adjustment; (7) attitude toward the future. His analysis indicates the following results:

Health and strength. A total of 134 stories, or 27.4 per cent of the themes, expressed attitudes toward health and strength. Of this number, 70.1 per cent expressed the attitude that health and strength are necessary for accomplishment, while 28.4 per cent feel that they are being deprived of normal contacts because of their physical handicaps.

Social adjustment. Of 365 stories, 73 per cent expressed attitudes toward social adjustment. Attitudes of bitterness and cynicism were most prevalent (45 per cent). Apprehensiveness was expressed in 38 per cent of the stories; and affectionate, congenial friendships were expressed in only 17.3 per cent.

Emotional maturity. Two hundred seven of the themes were rated under this category. The tendency to act impulsively was most prevalent (42.5 per cent). Mature emotionality appeared in only 13.5 per cent of the themes.

Emotional tone. Four hundred thirty stories were rated under this category. Feelings of insecurity were revealed most frequently (32.2 per cent). Depression was evidenced in 112 of the themes and is one of the most outstanding features of the personality structure. Happiness is shown in only 10 per cent of the stories.

Role in life. Of 430 stories rated in this category, ambivalence and indecision of action were the most prominent manifestations (41.9 per cent). Submission and dependence were the next most prominent expressions (36.3 per cent).

Attitude toward sexual adjustment. Only 92, or 18.4 per cent, of the stories deal with this category. Of this number, 59.8 per cent indicated maladjustment in the psychosexual sphere. Ambivalence was expressed in 12 per cent of the stories, and good adjustment in 28.2 per cent.

Attitude toward future. Expressions toward the future were expressed in 442 stories, or 88.4 per cent of the total number. The most outstanding characteristic was an attitude of depression (43.4 per cent). A type of autistic thinking or unrealistic success was the next most prevalent attitude toward the future (25.8 per cent). Fear of the future was expressed in 16.1 per cent, while hopeful accomplishment was expressed in only 14.7 per cent.

Summarizing the personality characteristics and attitudes of the paraplegic patients in this study, we find: (1) Physical strength and health are considered necessary for confidence and accomplishment, and that the patients generally have not accepted their injury and its limitations. Conflict between desires and inability to perform permeates the personality and produces a feeling of dependency. (2) There is ambivalence concerning role in life, producing indecisiveness and feelings of insecurity, which are reflected in social adjustment and relative lack of emotional control. Feelings of apprehension and cynicism are manifested in the sphere of social adjustment. (3) There is a relative lack of conscious expression concerning sexual adjustment. However, analysis of obtained data reveals maladjustment in the psychosexual sphere. (4) Depression is the most prevalent personality characteristic. It is revealed in the general tone of expression as well as in the outlook toward the future. Unrealistic success, a form of autistic thinking, also characterizes the outlook toward the future.

These basic psychological factors should be

considered in the rehabilitation of paraplegic patients. Mueller and Thompson studied an unselected group of 140 paraplegic patients, 19 of whom were partial quadriplegics, in order to appraise the effectiveness of psychological evaluation as a means of motivating rehabilitation activities. Each patient was given tests of intelligence and vocational interests. In addition, he had numerous guidance interviews with physicians, psychologists, vocational guidance counselors, social workers, and rehabilitation workers. The importance of the role that psychological factors play in rehabilitation was recognized. It should be emphasized that a tremendous psychological load is placed upon the personality as it must absorb the traumatic effects of injury, shock, extended periods of illness, severe physical handicaps involving loss of nerve and muscle control, impaired bowel and bladder fuctioning, altered sex functioning, and the multitude of adjustments that must be made in facing everyday problems of life. Considerable situational depression occurs, especially in the early stage of paraplegia. There are feelings of hopelessness, resentment, self-pity, dependence, and other abnormal attitudes that require psychological treatment in order that the patient may face his problems realistically and will be willing to take an active part in rehabilitating himself.

The studies here reported agree that depressive reaction is characteristic of many paraplegic patients, and that they have a tendency to react with less mature emotional expression, to show greater dependency needs, and to be unrealistic in their thinking. The suggested solutions for overcoming these emotional handicaps are largely psychological in nature and involve helping the patient accept reality, establish independence and self-sufficiency, and take an active part in his own rehabilitation.

This was essentially the procedure employed by Mueller and Thompson. By means of psychological evaluation, motivation, stimulation, and morale building, the patient was helped to overcome his frustrations and to establish healthy emotional and attitudinal adjustments to his disability. The program involved gaining the patient's confidence and cooperation, overcoming wishful thinking regarding recovery, occupational plans, and home and community living, and to view them realistically and purposefully, to the end that he would participate willingly in his rehabilitation program. The responsibility for motivation was a cooperative undertaking shared by the entire staff. A certain amount of initial resistance was found in many patients to rehabilitation activities requiring physical and mental effort on their part. Characteristic reactions were: indifferent, ambivalent, or depressive attitudes toward the future and the possibilities it could hold for them socially and vocationally; indecisiveness and feelings of insecurity; and assuming a submissive and dependent role in life, with accompanying immature emotional behavior.

In order to change these unhealthy mental states and resolve emotional conflicts, a concerted program of motivation was undertaken aimed at mobilizing physical, mental, and personality resources in the patient and directing them toward his rehabilitation. To achieve maximum rehabilitation of patients with spinal cord injuries, we must consider not only restoration of maximal physical and emotional fitness, but also training for vocational and occupational activities commensurate with mental abilities, vocational interests and aptitudes, and personality characteristics.

The rehabilitation program was comprehensive in order that the needs, interests and capacities of individual patients could be satisfied. Throughout the rehabilitation process, stress was placed on individual therapy. Activities were not engaged in for the sake of activity, per se, but were evaluated in terms of their therapeutic significance. There was an interplay of both physical and psychological medicine, stressing treatment of the patient as a whole, physically and emotionally. As the patient regains physical, social, and vocational independence through this therapeutic process, he gradually regains confidence, overcomes fear, loses the attitude of hopelessness and defeatism, and reflects a fundamental change in attitude and outlook.

The experimental study shows that the program is producing significant results in terms of personal, social, and vocational adjustment. One hundred thirteen, or 80 per cent, of the patients included in the study are actively pursuing their expressed vocational interests. Fifty have been discharged from the hospital during the one-year period of the study.

The psychological thread that runs through all rehabilitation activities is that of therapy through interest and motivation. Without these, rehabilitation activities are empty shells and the process is doomed to failure. Recognizing their importance, special efforts were made to discover the capacities, interests, and personality characteristics of each patient and to use them as foundations upon which to build a sound rehabilitation program. Psychological tests were used both for diagnostic and motivational purposes. Information revealed by the tests was invaluable in assisting physicians and rehabilitation staff workers in evaluating and understanding patients, as well as in helping patients to know and understand themselves better. With this knowledge, the way is opened for a sound cooperative rehabilitation program in which the patient's personality becomes the center of the therapeutic process, and the activity becomes the means for carrying on therapy. The program is eminently successful as judged by an estimate of upwards of 70 per cent who achieve good psychological adjustment, and the high percentage of patients—approximately 80 per cent—who have been discharged from the hospital and who are successfully adjusting to normal community life at home.

SUMMARY AND CONCLUSION

Personality problems of the spinal cord injured, as revealed by recent psychological studies, are summarized, and their bearing on the rehabilitation of these patients is outlined. Personality characteristics and attitudes revealed most frequently are: situational depression associated with the disability and with future role in life; a tendency to immature emotional expression characterized by impulsiveness, explosiveness, and egocentric behavior; a tendency toward ambivalence and indecision characterized by passive, submissive, and dependent behavior; and, a tendency toward autistic or unrealistic thinking regarding problems of life.

The suggested solutions for overcoming these emotional handicaps are largely psycholgical in nature, and involve helping the patient, through various therapies, to accept himself and his problems on a mature emotional basis, assume a realistic attitude toward his future activities, establish independence and self-sufficiency, and take an active part in his own rehabilitation.

A successful program of rehabilitation, based on a psychological approach, is discussed.

Received July 25, 1949.

REFERENCES

1. McLAUGHLIN, E. J. An investigation of the personality characteristics and attitudes of the paraplegic patient. Unpublished M.A. thesis, Fordham Univ., 1948.
2. MOSAK, H. H. Personality adjustment of paraplegic veterans. Hines, Ill.: Hines Paraplegic Center, 1948.
3. MUELLER, A. D., AND THOMPSON, C. E. Psychological aspects of the problems in spinal cord injuries. Memphis, Tenn.: Kennedy Veterans Administration Medical Teaching Group, 1949.

THE MEASUREMENT OF INTELLIGENCE OF ONE HUNDRED TWO MALE PARAPLEGICS[1]

MORSE P. MANSON
BIRMINGHAM VETERANS ADMINISTRATION HOSPITAL
VAN NUYS, CALIFORNIA

APPROXIMATELY one thousand healthy young adults each year in the United States are made helpless by severe spinal cord injuries [5]. Therefore, the rehabilitation of the paraplegic may be considered to be a permanent problem.

Psychologically, the paraplegic absorbs the trauma of pain, shock, infection, extensive and intensive hospitalization, loss of movement, reduced — if not totally lost — sexual function, impaired bladder and bowel control and feelings of depression and dependence. With a single blow to the spinal cord an individual, perhaps at the peak of his physical development, is reduced from the independence of an adult to the dependence of a neonate. No personality, no matter how strong and stable, can face such a devastating onslaught without severe psychic repercussions [2, 3].

Because of the severity of the physical and psychological injuries to the paraplegic, many rehabilitation workers have felt that there might be serious intellectual impairment. Psychiatrists often state that there is a "breaking point" for each person, implying that to go beyond this point would result in a breakdown of emotional and mental functions. Do the trauma of paraplegia take the paraplegic beyond the "breaking point"? Are there *any* changes in intelligence attributable to the injuries accompanying paraplegia? An analysis of intelligence test scores of paraplegics may answer these questions.

No studies dealing with the measurement of intelligence of paraplegics have been seen in the literature. One paper [7] reported the use of projective techniques on paraplegics but found such techniques to have less value than the interview; besides, paraplegics showed no disposition to participate in any tests which did not seem to have immediate value to them. Two studies mention the results of mental ability tests taken *prior* to the injury. Thom *et al.* [7] noted that general mental ability ranged from low scores in Grade IV to high scores in Grade I on the Army General Classification test. On the same test, Ginzberg and Klarman [1] found paraplegics to be about equally divided among "low," "medium," and "high" scores. It thus seems that there might have been a normal distribution of scores on the AGCT prior to paraplegia.

METHOD

Practically all the paraplegics in Birmingham Veterans Administration Hospital had been interviewed concerning their work plans upon discharge after reaching maximum hospital benefits. The various resources of the hospital rehabilitation program were described and the paraplegics were urged to use these services. Vocational advisement was but one of many services. Some paraplegics did not feel the need for advisement and refused to participate, stating that work of any kind would be impossible for them. These had decided to live upon their pensions. Others had plans which they hoped to follow without assistance from any one; still others had plans but felt, for varied reasons, not yet ready to take positive action. However, a large percentage were interested in vocational advisement and planning. Many of these were given intelligence tests as well as other tests. These intelligence tests supplied the data in this study.

It has been observed [1] that "those who had the greatest schooling were most desirous

[1]Sponsored by the Veterans Administration and published with the approval of the Chief Medical Director. The statements and conclusions published by the author are a result of his own study and do not necessarily reflect the opinion or policy of the Veterans Administration.

of further schooling, the most favorable disposition being found at the college level, where four out of five wanted to continue their education. . .about 70 per cent of those who did not desire further schooling were also disinterested in occupational training." It seems reasonable to assume that the group using advisement services was better motivated, had more education, and was more intelligent than the group not using advisement services.

It is thus probable that the group of paraplegics availing itself of vocational advisement services was a select one and not representative of the total paraplegic population in Veterans Administration hospitals.

The intelligence scores of 102 paraplegic veterans, all male, were obtained and analyzed. Within this group, 84 paraplegics had completed vocational advisement and developed training programs leading to specific employment objectives. 18 paraplegics, at the time of this study, had not completed advisement. These two subgroups were compared.

GROUPS

The age range for the total group was from 18 to 38 years, with a mean age of 25.3 years. The subgroup completing advisement was slightly younger than the subgroup not completing advisement. The educational range of years of education completed was from 6 to 15 years for the completed subgroup and 3 to 15 years for the incompleted subgroup; the means of years of education completed were 11.1 and 10.2 years respectively. A difference significant at the five per cent level of confidence was found for years of education completed.

In the subgroup completing advisement, 55 per cent had been privates or privates first class, 28 per cent noncommissioned officers and 17 per cent commissioned officers. The percentage of commissioned officers is much larger than would be found in a random sample, once again indicating that this group of paraplegics is a select one. Eighty-four per cent of completed paraplegics had been in the Army, 8 per cent in the Navy and 8 per cent in the Marine Corps. Of the subgroup completing advisement 86 per cent were service connected disability cases compared with 76 per cent of the subgroup not completing advisement; thus 12 per cent and 24 per cent of the respective subgroups were non-service-connected paraplegics, being accounted for by civilian-type accidents. Approximately 90 per cent of the total group were white veterans and 10 per cent nonwhite veterans.

RESULTS

Three different intelligence tests were used. They were: (a) California Test of Mental Maturity, Advanced '47 S-Form [6], (b) Otis Quick-Scoring Mental Ability Test, Gamma, Form Am [4], and (c) Wechsler-Bellevue Scale, Form I, [8].

On each of the three tests, as well as on a composite mean of the three tests, the mean intelligence quotient for the group of 102 paraplegics was above 100. On two mean scores, for the nonlanguage factors of the California for the incompleted subgroup, and the Otis for the incompleted subgroup, did the mean IQ fall below 100.

The mean IQ for the total group on the California was 108.9, placing at the 70th percentile. The completed subgroup made a higher mean IQ than the incompleted subgroup, but the critical ratio of 1.4, although indicating a trend is not statistically significant. A mean IQ of 114.1 on the language subtests was found for the total group. The difference between the subgroups on language IQ's was insignificant. The mean IQ on nonlanguage subtests for the total group was 106.1. A highly significant difference was found between the completed and incompleted subgroups, indicated by a critical ratio of 4.3, on nonlanguage test performance.

If it is assumed that the language tests on the California measure relatively stable mental functions while the nonlanguage tests are more prone to change with mental deterioration, a measure of deterioration may be constructed. This crude measure might be indicated by dividing the nonlanguage IQ by the language IQ. If the nonlanguage IQ is the same or higher than the language IQ it would be assumed that no mental deterioration had occurred.

Using the above formula, 4.8 per cent deterioration was found for the paraplegics completing advisement; 26.2 per cent for those not completing advisement; and 7.1 per cent for the total group taking the California test. A percentage higher than 10 per cent might be

TABLE 1
CALIFORNIA TEST OF MENTAL MATURITY

Groups	No.	Range of IQ's	Mean IQ	SD of IQ's	Critical Ratio	Deterioration %
			Language subtests			
Completed Advisement	57	64-156	114.0	20.0		4.8
Incompleted Advisement	11	69-141	114.5	23.0	.1	26.2
Both Groups	68	64-156	114.1	20.8		7.1
			Non-language subtests			
Completed Advisement	57	66-138	108.5	9.8		
Incompleted Advisement	11	80-130	84.5	16.8	4.3	
Both Groups	68	66-138	106.1	10.6		
			All subtests			
Completed Advisement	57	75-141	112.9	14.0		
Incompleted Advisement	11	75-125	105.3	15.5	1.4	
Both Groups	68	75-125	108.9	15.2		

viewed with suspicion of mental deterioration (Table 1).

The mean IQ for the group of 23 paraplegics taking the Otis intelligence test was 106.2. There was no significant difference between the subgroups, although the subgroup completing advisement made the higher mean IQ (Table 2).

TABLE 2
OTIS QUICK-SCORING TEST OF MENTAL MATURITY

Groups	No.	Range of IQ'S	Mean IQ	SD of IQ s	Critical Ratio
Completed Advisement	18	80-129	108.4	14.2	
Incompleted Advisement	5	67-123	89.5	13.5	1.3
Both groups	23	67-129	106.2	16.6	

Eleven paraplegics were administered the Wechsler-Bellevue test of intelligence. The mean IQ on the total test was 112.7; the completed group, once again, made a higher mean IQ than the incompleted group, but the difference was not statistically significant. No significant differences were found when the two subgroups were compared on the verbal or the performance subtests, although the completed subgroup made higher mean IQ's on the verbal and performance tests.

Since the individual subtest scores were not available, making it impossible to determine the extent of mental deterioration as defined by Wechsler [9], the same crude determination of deterioration—as made in the California test—was made, and no deterioration was

TABLE 3
WECHSLER-BELLEVUE TEST OF INTELLIGENCE

Groups	No.	Range of IQ's	Mean IQ	SD of IQ's	Critical Ratio	Deterioration %
			Verbal subtests			
Completed Adv.	9	86-140	109.0	18.0		0
Incompleted	2	89-113	101.0	16.9	.4	0
Both Groups	11	86-140	107.2	18.1		0
			Performance subtests			
Completed Adv.	9	104-129	117.0	10.3		
Incompleted	2	93-112	102.5	13.4	1.1	
Both Groups	11	93-129	113.6	11.7		
			Total subtests			
Completed	9	95-136	114.5	14.1		
Incompleted	2	93-114	103.5	14.9	.7	
Both Groups	11	93-136	112.7	14.0		

noted. However, on two of the paraplegics completing advisement, the Wechsler subtest scores were available. One paraplegic had 8 per cent and the other had 10 per cent deterioration. The scant evidence of this study would seem to indicate that the paraplegics tested as a group did not reveal mental deterioration to any marked extent, if at all (Table 3).

When the results of the three intelligence tests were combined, a total mean IQ for the group of 102 paraplegics was 108.7. The 84 paraplegics completing advisement were higher in mean IQ than the 18 paraplegics not completing advisement, but the difference was not significant (Table 4).

TABLE 4
COMBINED INTELLIGENCE TEST RESULTS

Groups	No.	Range of IQ'S	Mean IQ	SD of IQ s	Critical Ratio
Completed Adv.	84	75-141	109.7	14.9	
Incompleted Adv.	18	67-125	104.0	17.3	.6
Both groups	102	67-141	108.7	15.5	

DISCUSSION

Because paraplegics receive severe physiological and psychological damage the question frequently is raised by rehabilitation workers if intellectual impairment has occurred. There is some slight evidence in previous studies that paraplegics were normally distributed in intelligence prior to their injuries. In the group of 102 paraplegics studied the mean IQ's on three different intelligence tests indicated that this group was of high average intelligence. Because this group appeared to be highly motivated, had fairly high average educational achievements, and had a fairly large percentage of commissioned and noncommissioned officers it is believed that the group was a select one and not representative of the paraplegic population in veteran's hospitals. It is very likely that groups of paraplegics not interested in vocational planning may have lower mean IQ's. However, the impression of the writer is that representative groups of paraplegics would not fall below the mean intelligence of the general population, but probably be slightly higher since selection factors had operated in the selection of military personnel.

There was no conclusive evidence that mental deterioration had occurred in the group studied. Only crude attempts had been made to determine this. Nevertheless the opinion, both from the available test results and from personal observations, is that paraplegia per se is not a precursor of mental impairment.

CONCLUSIONS

1. A group of 102 male paraplegic veterans had a mean IQ of 108.7. Paraplegics completing vocational advisement and developing training programs leading to specific work objectives had higher mean IQ's than paraplegics not completing advisement. But the difference in mean IQ's was not statistically significant.

2. No substantial evidence of mental deterioration due to injuries pertaining to paraplegia was found.

Received August 8, 1949.

REFERENCES

1. GINZBERG, E. AND KLARMAN, H. The paraplegic. *Bull. U. S. Army med. Dep.*, 1947, **7**, 892-897.
2. MANSON, M. P. The concept of rehabilitation applied to paraplegia. *Arch. industr. Hyg. & occup. Med.*, 1950, **1**, 65-72.
3. MANSON, M. P. Psychological aspects in the rehabilitation of paraplegics (unpublished).
4. OTIS, A. S. Otis quick-scoring mental ability test, Gamma test, Form Am. Yonkers-on-Hudson, N. Y.: World Book Co., 1937.
5. SPURLING, R. G. Neurosurgery (and the neurosurgeon) in World War II. *J. Amer. Med. Ass.*, 1947, **135**, 473-476.
6. SULLIVAN, ELIZABETH T., CLARK, W. W., AND TIEGS, E. W. California short-form test of mental maturity, Advanced '47 S-Form. Los Angeles, Calif.: California Test Bureau, 1947.
7. THOM, D. A., VONSALZEN, C. F., AND FROMME, A. Psychological aspects of the paraplegic patient. *Med. Clin. N. Amer.*, 1946, **30**, 473-480.
8. WECHSLER, D. Wechsler-Bellevue intelligence scale, Form I. New York: Psychological Corp., 1939.
9. WECHSLER, D. *The measurement of adult intelligence.* (3rd Ed.) Baltimore, Md.: Williams & Wilkins, 1944.

THE EFFECT OF FOCAL BRAIN LESIONS ON RECALL AND ON THE PRODUCTION OF ROTATIONS IN THE BENDER GESTALT TEST[1]

LEO J. HANVIK AND A. LLOYD ANDERSEN
MINNEAPOLIS VETERANS ADMINISTRATION HOSPITAL

THIS report is concerned with the Bender Gestalt protocols of a group of brain-damaged patients in each of whom the locus of the cerebral lesion had been made known by surgery or through X-ray verification. The total group of 44 patients was made up of 20 patients with lesions in the "dominant" hemisphere (presumed to be the left hemisphere in right-handed persons) and 24 patients with lesions in the other, or "nondominant," hemisphere. All were male patients in the Minneapolis Veterans Administration Hospital and all had been subjected to fairly exhaustive psychological study which included, among other things, the administration of the Bender Gestalt test.

PROCEDURE

The Bender test was administered to all patients in the standard manner [1]. In addition to having the patients copy the nine drawings from the cards, the examiners also tested the ability of each patient to draw the gestalten from memory, i.e., without looking at the model drawings. The protocols were then examined with a view to noting what differences, if any, obtained between the drawings of the group of patients with dominant-hemisphere lesions and those of the group with nondominant-hemispherel lesions.

In order to compare the two groups, two measures capable of quantification were obtained: (1) the mean number of figures recalled correctly by each patient subgroup, and (2) the per cent of cases in which at least one drawing had been distorted by rotating it 30 degrees or more from the original plane. These two variables were chosen, not only because a quantitative score could be obtained on them, but also because extensive clinical use of the Bender test had suggested that the tendency to rotate drawings together with a tendency toward poor recall of the figures might be particularly characteristic of brain-damaged patients, and that patients with brain lesions in different locations might differ with respect to these variables.

To extend the study a little further, a control group of other hospital patients was used. This control group consisted of patients whose presenting complaint on admission to the hospital was low back pain. These patients had been studied as a part of a different investigation [2] and had been found finally to be made up actually of two clinical types: (a) fourteen patients with herniated intervertebral disc, (ultimately surgically verified) and (b) twenty-three patients whose general physical and neurological examinations had been essentially negative and in whom the low back syndrome was deemed to be psychogenic in origin.

As was done in comparing the two subgroups of brain-damaged patients, the two subgroups of controls were compared with each other with respect to per cent of cases rotating and with respect to recall. The control subgroups did not differ significantly from one another on these two variables. In recall score, the total control group was approximately the same as a large sample of normals studied by Welsh (Unpublished research). Data were not available on rotations for a normal group. Also, comparisons on the two variables were made

[1]Published with the permission of the Chief Medical Director, Department of Medicine and Surgery, Veterans Administration, who assumes no responsibility for the opinions expressed nor conclusions drawn by the authors. Gratitude is expressed for the aid and cooperation of the Neurology staff at the Minneapolis Veterans Administration Hospital.

between the nondominant-hemisphere group and the control group as a whole, and between the dominant-hemisphere group and the control group. The results of the latter comparisons together with the significance of the findings, are shown in Tables 1 and 2.

TABLE 1
PERCENTAGE OF CASES PRODUCING ONE OR MORE ROTATIONS IN THE BENDER-GESTALT DRAWINGS: RESULTS OF COMPARISONS

	% Brain Damaged	% Controls	Difference	CR	P
Dominant-Hemisphere Damage:	55.00	18.91	36.09	2.81	<.010
Nondominant-Hemisphere Damage:	62.50	18.91	43.59	3.70	<.001
Total	59.09	18.91	40.18	4.09	<.001

TABLE 2
MEAN NUMBER OF BENDER GESTALT FIGURES REPRODUCED ON RECALL

	Brain-Damaged Patients Mean	SD	Control Group Mean	SD	Difference	CR	P
Dominant-Hemisphere Damage:	3.90	1.97	5.57	1.73	1.67	.64	>.05
Nondominant-Hemisphere Damage:	4.67	1.58	5.57	1.73	.90	.38	>.05
Total	4.32	1.81	5.57	1.73	1.25	.50	>.05

RESULTS

1. The dominant-hemisphere group is statistically not significantly different from the nondominant-hemisphere group with respect to either the number of figures reproduced on the recall exercise or percentage of cases rotating the figures.
2. There are no significant differences between either of the brain-damaged subgroups and the control group as to number of figures recalled.
3. There *are* significant differences (P. is less than .01) between the brain-damaged group and the control group on percentage of cases rotating the figures. The brain damaged group has significantly more instances in which at least one figure is rotated in the protocol.

Other facts of interest which came to light in these comparisons (and which are *trends* rather than statistically significant differences), are these:

A. Rotations occur in *both* "normals" (e.g. the patients with herniated disc) and in "neurotics" (e.g. the cases of probable functional low back pain), as well as in patients with brain damage.

B. There is a slight trend toward poorer recall scores among patients with dominant-hemisphere lesions, in comparison with patients with nondominant-hemisphere lesions.

C. There is a similar slight trend among the nondominant-hemisphere-damaged patients to be more prone to rotate the figures than is the case among dominant-hemisphere-damaged patients.

SUMMARY

Statistics are presented relative to the Bender recall scores, and the percentage of cases rotating the Bender figures, when the Bender protocols of a sample of brain-damaged patients are studied. Comparisons are made between patients with lesions in the dominant hemisphere and those with lesions in the nondominant hemisphere. Comparisons are also made between the recall scores and rotative tendencies of the brain-damaged patients and measures of the same variables for a control group. The only difference appearing which is statistically significant is the difference between the brain-damaged group and control group on percentage of cases rotating the Bender figures. The brain-damaged group exceeds the control group significantly in this characteristic.

Received July 18, 1949.

REFERENCES

1. BENDER, LAURETTA. *A visual motor gestalt test and its clinical use.* New York: Amer. Orthopsychiat. Ass., 1938.
2. HANVIK, L. J. Some psychological dimensions of low back pain. Unpublished Ph.D. thesis, University of Minnesota, 1949.

A BASIS FOR INTEGRATION RATHER THAN FRAGMENTATION IN PSYCHOTHERAPY

REX MADISON COLLIER
UNIVERSITY OF ILLINOIS

SINCE the appearance of Rogers' *Counseling and psychotherapy* [9] in 1942, we have observed a very rapid increase in the frequency with which psychologists have given attention to both the theoretical and the practical problems of psychotherapy. The publication of *Counseling and psychotherapy* was, of course, coincidental with the increase in demand for psychotherapy from clinical psychologists in the military assignments. The book not only gave the clinical counselor or psychotherapist a technique and a code, but also presented the first verbatim transcript of a series of recorded therapy interviews. To many psychologists, it has seemed that, at last, we have stated for ourselves a point of view and a technique that is not only distinctly psychological but is also completely legitimate for use by the psychologist. The next tempting step was easily taken, namely, to emphasize contrasts between "traditional" psychotherapies and the "new" psychotherapy.

While the purpose of this paper is not to evaluate the significance of Rogers' "nondirective" or "client-centered" emphases, reference must be made to its influence in current psychotherapy practice. The purpose of the paper is rather to indicate the fragmenting tendencies in the field of psychotherapy and to suggest an approach which may aid in constructive integration. Divergent emphases often occur in a field of science prior to greater unification. The history of psychology could illustrate a number of these sequences of (a) fragmentation, (b) a degree of resolution of the controversies, and (c) new trends. Twenty-five years ago there appeared *The psychologies of 1925* [7] and subsequently *The psychologies of 1930* [8]. The divergence of theory consumed considerable time and attention then. However, while there are still differences of emphasis, it would be far less appropriate currently to extend this series. The situation in psychotherapy may very well run a kind of parallel. It would be possible and even desirable to publish a *Psychotherapies of 1950*; there might well be another such volume in 1955, but eventually integrating approaches should occur which would reduce current confusion and increase certainty.

The fact that all therapies have successes and all therapies have failures strongly suggests that there should be bases of integration. No therapy technique can be credited with either all the "cures" or all the failures. That some therapists function more comfortably and effectively with one emphasis rather than another suggests that personality differences among therapists may be as important as differences in technique. Clues for integrating concepts are not likely to be found, however, by studying personality differences of therapists. A notable current movement to unify theory and technique of psychotherapy is spearheaded by the work of Mowrer [6] in his attempt to relate psychotherapy to learning theory.

The thesis of this paper is that bases for greater unity should be found in the discovery of continuities among the common aspects of therapy some of which are emphasized by one school, others emphasized by other schools. Without attempting an exhaustive list, these common aspects may be partially detailed as follows:

1. Techniques
   a. Interpretation
   b. Clarification and reflection of feeling
   c. Reassurance
   d. Pressure and persuasion
   e. Questioning
   f. Free association

g. Hypnotic procedures
h. Drugs

2. Process and content[1] (Typical categories for classifying the actual verbal exchanges between therapist and client.)
   a. Understanding or insight
   b. Discussion of plans
   c. Answers to questions
   d. Rejection or acceptance of interpretation
3. Relationship (Reference is made here to the relationship between therapist and client which may in part be described by the dimensions indicated.)
   a. Balance of responsibility between therapist and client or patient.
   b. Degree of explicit and implicit freedom of response (limits) for patient.
   c. Degree of mutual empathies.
   d. Degree of mutual acceptance and rejection.
   e. Degree of transference and countertransference.

Overlapping among the dimensions listed under 3 is inevitable; or, more quantitatively, most such dimensions would have tendencies to converge. It is not the purpose of this paper to pursue that aspect of the problem.

4. Criteria of change[2] (Indicators of improvement)
   a. DRQ as a measure of change in psychological tension [4]
   b. Insight
   c. Increased self-acceptance
   d. Reduction in hostilities
   e. Increases in range of social activities and reduction of old neurotic restrictions
   f. Increases in reality thinking and planning.

Disagreements and special emphases have arisen principally with regard to concepts of technique and relationship. Bases of integration can, therefore, best be illustrated by reference to these two areas. The continuum concept[3] is thus far most easily illustrated with reference to the relationship between therapist and patient. It is not difficult, for example, to see how the balance of responsibility for attacking and handling the problem can vary between the therapist and the patient. Likewise, it is not difficult to see that the other aspects of relationship may vary according to degree or dimension. It is not quite so obvious how an aspect of technique may vary directionally and become the basis of resolving a dichotomy.

THE INTERPRETATION CONTINUUM

The Rogerian school has often contrasted interpretation with "reflection of feeling." The one is referred to as directive; the other as nondirective. The hypothesis here proposed is that, if we can assume a continuum described as *the degree of uncovering or revelatory interpretation,* then "reflection of feeling" will become a segment of the more comprehensive dimension. "Reflection of feeling" would, therefore, be a form of interpretation and could be rated in terms of the degree to which it appears to reveal feelings, tendencies, aspects of personality heretofore unrecognized. Demonstration of this approach would be a step toward eliminating both an unneeded dichotomy and undesirable misrepresentations of psychoanalytic procedure. Let the continuum be represented by the line with terminal points *a* and *b* (Figure 1).

FIGURE 1

a would represent the least degree of uncovering interpretation while *b* would represent the maximum degree of the variable. What seems to be steps along this continuum may be illustrated by excerpts from verbatim published recordings.[4] Exactitude in identifying posi-

[1]Cf. Snyder, W. U. [14] for more complete description and listing; cf. also Seeman [13]. Subcategories are illustrative without purporting completeness.

[2]See Curran [2]; cf. also Snyder [14] and Seeman [13].

[3]cf. E. S. BORDIN, Dimensions of the counseling process. *J. clin. Psychol.*, 1948, 4, 240-244. This significant article, presenting a somewhat similar line of thought, came to the writer's attention after the present paper had been accepted for publication. Extended discussion, therefore, cannot be made at this time.

[4]Experimental verification of these classificatory steps can be made by the method of Thurstone and Chave [16] in determining scale values for an attitude scale.

tions on the scale or continuum is not the present purpose, but rather to indicate that continuity exists and the probability of experimental definition of position.

The least degree of uncovering interpretation or the extreme left of the continuum would be simple repetition of the subject's response. Examples of this may be found in Axline's *Play therapy* [1, p. 103].

a. Jack: Look, I spit on my home.[5]
Therapist: You spit on your home.

b. Jack: I hate my father! I hate my brother!
Therapist: You hate your father! You hate your brother.

While the therapist's purpose here includes acceptance of the subject (note that several continua may be involved in a single exchange between therapist and subject) there is the probability that the subject has been made slightly more aware of his own feelings and impulses. Since the technique is designed, in addition to being acceptive, to make the subject more aware of impulses and feelings which heretofore had escaped his recognition, the intent of the therapist is revelation even though it may be only in a mild degree. For purposes of this paper, the above example will be called Step I on this continuum.

Some degree of advance, although probably small, in the direction of the opposite end of the continuum is represented by selecting a portion of the subject's response and repeating it or restating it. The therapist uncovers here in the sense of pointing out or calling attention to what seems to him the more significant aspect of the subject's response. Examples of Step II are as follows:

Richard: (who is going to see his mother) I'll get her something nice. I'll surprise her with a nice present.
Therapist: You'll take her a nice surprise.
Richard: Yes, I'll be gone two weeks. Maybe three. Gosh, it'll be good getting out of here.
Therapist: You'll be glad to get away from here for a while.
Richard: I'll be ten years old. I'm in the fourth grade. My dad wants me in the fifth grade, but I tell him I don't want to be in the fifth grade. I like to fail.
Therapist: You like to fail. [1, p. 123]

[5]While attention is directed to the response of the therapist, the subject's previous response must be included for context.

These responses of the therapist are not simple repetitions, although some of them approach the repetitive. The therapist has selected and restated. In the last exchange, the therapist could have selected other aspects of Richard's response, e.g.,

a. You'll be ten years old.
b. You're in the fourth grade.
c. Your dad wants you in the fifth grade.
d. You tell your dad you don't want to be in the fifth grade.

The actual selection of the therapist is designed (in addition to being acceptive and permissive) to aid in recognition of a significant attitude. There is, therefore, a greater degree of revelation than in simple repetition.

A still further step (III) in the direction of uncovering the significant, latent or unverbalized aspects of the subject's response is to restate what the subject has said for the purpose of increased clarity or for the purpose of revealing (reflecting?) an affective tone or trend. The increase in clarity is directed toward what the therapist perceived to be an improved revelation of the problem as *currently* represented. This is somewhat more than selecting an aspect of the subject's response; it restates, summarizes or epitomizes for the purpose of revealing more clearly to the subject the core or implications of what he has said. Therapists can restrict themselves to Steps I and II and avoid the need for any very penetrating inferences, but beginning with Step III inferences that penetrate beneath the surface become increasingly apparent. Examples illustrating Step III are as follows:

a. Subject: . . . . I never used to worry about people or what they thought. I didn't mind what they thought about me or my personal affairs, until this last year. I didn't think I was so bad until then. In fact, at that time I thought I was pretty darn goodlooking. Then I began to look at fellows and think, "Now if I had John's nose, and Frank's eyebrows, and Bud's hair, and Harry's chin, I really would look like something. Then I'd be a swell fellow."
Counselor: You sort of wanted to be the best of everything. [15, p. 104]

- - - - - - - - - - - - - - - - - - -

b. Subject: . . . . The only thing is that I can't help but feel that I work against myself. I mean it's just so obvious to me that I *do* work against myself. And it seems rather silly to be one minute trying to do something about it, and

then it seems so certain that there's something present that doesn't want to do anything about it. (*Pause*) It's such a contrary thing — the whole problem is just - - -

Counselor: So that the very time that one part of you is perhaps really taking hold of this whole situation, another part of you is just sabotaging the whole business. [15, p. 173]

Another step (IV) in the direction of a greater degree of revelatory interpretation is represented when the therapist combines previous statements of the subject with current statements for the purpose of stimulating integrative insights. The uncovering aspect of the therapist's response is evident in his attempt to bring to a clearer conscious level personality trends that heretofore had been unrecognized. The published verbatim material representing "nondirective" technique shows few, if any, examples of this. Psychoanalytic material would contain a high frequency of this type of therapist response. Analyst responses, however, do not begin at this point, for they may be found to spread over the whole range of this continuum. Examples of Step IV are as follows:

Dr. Do you remember that you had the same feeling in connection with mother, for whose health you were concerned, and that you behaved as if you were she? Do you remember that I reminded you of a statement of little Ralph . . . .

Pt. I do

Dr. . . . . who once said "If I don't stop behaving so badly, it will kill her." Do you remember you said it?

Pt. That's right.

Dr. And that you had to run home to see whether she were still alive and that you had a bad conscience as if you had done something wrong? Then you said something wrong means: something is wrong with your heart or you did something wrong. (Pause). Now, you said when your girl friend complained about her heart, you felt as if you were she. [3, pp. 132, 133]

In suggesting and illustrating hypothetical Step V, it should perhaps be pointed out again that the writer is not here attempting to indicate with any degree of precision these positions. That remains for experimental determination. Furthermore, we must avoid the implication at this time of equal distances between these reference points. There is probably greater distance, i.e., psychological distance in terms of judges' ratings, between III and IV and between IV and V than between I and II or even between I and III. The writer's' hypothetical conception of this continuum in terms of the foregoing and subsequent illustrations is shown in Figure 2.

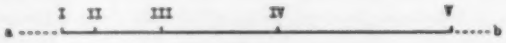

FIGURE 2

The dotted extensions simply mean that the limits have not yet been explored nor defined. For example, at *a* there could conceivably be a mechanical repetition of precisely what the patient had previously said. Experimental exploration of the effectiveness of this therapy has not been made, but it is a plausible procedure. At the other extreme *b*, the work of Rosen [12] suggests that the dimension in extendable beyond even traditional analytic technique.

The region of Step V is characterized by a deeper uncovering type of interpretation. The more elemental and the more deeply unconscious tendencies of the personality are here the objective of the therapist's techniques. The uncovering of oral fixations, castration complexes, strongly rejected incest or homosexual tendencies represent the region of Step V. Verbatim illustration of therapists' procedures and verbalizations are difficult to find since analysts have only just begun to record their interviews. The following two excerpts from apparently verbatim material illustrate approximations to Step V. The first example would probably be classified somewhat to the left while the second might be somewhat closer to this last position.

a. Dr. Without this anxiety, you probably wouldn't have heeded your aunt's warning, and you probably would have had sexual intercourse with Carl. From this point of view, the anxiety was a valuable help, because your sexual desire had destroyed your ability to resist your cousin. If your aunt's warning had appeared without anxiety, you probably would not have had the power to interrupt your sexual play and to escape unharmed. That was the situation at the time. Today, however, sexual intercourse is no longer a danger. Now you may conceive a child without being blamed for it. But you still behave as if you were a little girl. Why? [5, pp. 59, 60]

- - - - - - - - - - - - - - - - - - -

b. His voice was tired, as he said—"I'm still afraid of cats. . . ."

I drew a long breath. "Perhaps there's a connection between this attempted suicide and your phobia. When you decided to die, you had to cross

the street to go to the river. Perhaps the crossing of a street may represent, in your Unconscious, the wish to leave home in order to die. The fear you feel when you try to cross the street, would then be a repetition of that childhood fear when you were attacked by a cat. It would serve as the warning —

"Don't do it again!" [Ibid, p. 227]

That such therapist responses can be classified along a continuum such as the one described above implies neither their validity regarding personality dynamics of the subject nor the appropriateness of the response. These questions are irrelevant to the immediate problem of identifying these continua and their relational characteristics. It will be indicated later, however, that research on the effectiveness of therapy can be closely related to the dimensional concepts described here. At the moment it is important to point out that acceptance of even the one continuum just described reduces the appropriateness of emphasizing sharp demarcations between "non directive" technique and analytic technique. With regard to the technique of interpretation identified by this continuum, there is rather a gradual transition from one to the other with no necessary implication that the therapist becomes highly directive even at the opposite end of the continuum. Rogerian reflection of feeling may be represented as a segment near the *a* end of the continuum. To insist on limiting oneself to a particular segment of a dimension is, of course, possible but does not serve to make the technique distinctive. Nor is it desirable in the interests of fundamental clarity to insist on such sharp dichotomies as Rogerians frequently do.

If we are truly "client-centered"—and no therapy worth recognition is centered toward any other objective—then we will reduce our emphases on whether we are directive or nondirective, analytic, Meyerian, Adlerian, eclectic or otherwise; we will tend less toward being protagonists of "doctrines" and give attention to the needs of the client or patient. The hypothesis here submitted is that the needs vary from one client or patient to another and from one stage of therapy to another. Furthermore, in terms of these basic dimensions, the problem becomes one of determining the segments of the dimensions which should be emphasized for particular kinds of patients or problems and at successive stages of therapy. Rather than being rigid about the pattern of segments to be used or emphasized because of doctrinaire orientation, therapy that is truly client-centered and not arbitrarily restricted should be sufficiently flexible so that adaptation to the immediate needs of the client may be made. This suggested approach should provide a basis for unification of trends rather than for greater diversification. Sharp boundaries and dichotomies tend to blind one to the contributions of those whose results may have fundamental relevance to our own. It is pertinent here, also, to point out that no segment of such a continuum becomes necessarily the exclusive property of any particular group.

THE RESPONSIBILITY-BALANCE CONTINUUM

The limits of this paper prevent the discussion and illustration of more than one dimension or continuum under each of the categories of technique and relationship. Reference is made now to the one described as "balance of responsibility between therapist and client" for attacking the problem. The assumption that this responsibility balance is continuously variable obviates certain current confusions that beset students in their training in psychotherapy. The writer suspects this confusion to involve even the experienced therapist at times. At one extreme the relationship is structured so that the patient is entirely responsible while at the other extreme the therapist becomes completely responsible. The latter is approximated when the therapist actively elicits a case history from a patient. The former may be observed near or at the termination of a therapy period when, with increased maturity and insight, the patient plans his future activities. Again, it becomes essentially a question of what segment of this continuum is appropriate for a particular case, for particular stages of therapy, and for special social fields[6]. Experimental procedures will eventually reveal the answers to such questions.

While it is obvious that the case and therapy stage variables should affect this continuum, the implications of the social field within which therapy takes place are sometimes overlooked.

[6]O. H. Mowrer, after reading this manuscript, has aptly suggested that the segment pattern would vary also during a particular therapy session.

Frames of reference are subtly derived from social fields and often complicate the problem of communication. One could approach the problem of therapy as essentially a problem in communication. It would then be necessary to concern oneself extensively with the frames of reference of both patient and therapist. These frames of reference are, of course, the psychological contexts out of which issue specific verbalizations. The determination of the degree and frequency of coincidence of the frames of reference of therapist and patient might be particularly enlightening along with measures of therapy effectiveness and progress.

Specific application of this point may be made to the procedure of structuring initial relationships in therapy. Much has been said and implied in the literature about proceeding so that your client feels that it is his responsibility to produce, that he doesn't shift responsibility to the therapist for action and that the therapist must not behave as though he could solve another person's problems for him. The appropriateness of this approach will be found to vary with the social field. Consider patients in a hospital to be given psychotherapy. The counselor might say at the first interview something like this: "I can't solve your problems for you. No person can solve another person's problems for him. It is really up to you. Now, I'll see you every day at ten o'clock and we can discuss anything you like. Is there anything you would like to talk about today?"[7] While this may be nondirective in the textbook sense, it is far too rigid an observance of rule to allow the therapist to begin by trying to understand the needs of the patient, why he came to the hospital and what he expects.

What is a typical frame of reference of a patient coming to a hospital? Traditionally, one goes to a doctor or to a hospital to have something done to or for him. He gets a surgical operation, a physical examination, recommendations concerning diet, activities, intake of medicines, etc. The patient comes to a hospital, generally, because his efforts have failed; he comes as a last resort; if he could have helped himself, he would have done so and saved his time, his job and stayed among family and friends. With such a frame of reference, it must be a devastating blow to be told by one who represents the hospital, "I can't help you; it's up to you." or some similar verbalization of nondirective structuring. From the patient's point of view this is tantamount to a rejection of him and his problem. The therapist is more concerned here with being "correct" than in understanding his patient.

Conflict about the directive-nondirective issue seems to be resolved by applying the concept of the responsibility-balance continuum. In the social field of the hospital, the therapist acts more appropriately if he adopts initially a mode of behavior and verbalization that represents a rather high degree of therapist responsibility. In most instances the writer has not found it difficult gradually to shift this balance more and more to the patient as therapy progresses and as the patient's strength and confidence increase. Change with therapy is more similar to the growth of a child toward greater maturity than simply being analogous. The wise parent provides the child with opportunities to assume those responsibilities that are commensurate with his current capabilities. The wise therapist should behave in the same way. Effective therapy like effective child training eventually arrives at maturity and independence of the individual. With the patient as with the child, too early an insistance on independence and self-direction is equivalent to rejection. With children, fear and hostility often result; with patients, it is not uncommon for them to reject the institution or the therapist, or both, as inadequate. One may suspect that both children and patients have some justification on their side.

With the concept of the responsibility-balance continuum, it becomes less appropriate to emphasize dichotomies of directive and nondirective for this aspect of the therapy relationship. The problem becomes rather what segment of the continuum is associated with therapy progress under specified conditions. This may be subjected to experiment following experimental definition of continuum positions. These positions could be illustrated by excerpts from verbatim recordings as was done in an earlier part of this paper. Such positions could also be experimentally identified by the Thurstone technique to which reference has also been made.

In a social field such as a college counseling

[7]Cf. Rogers [10].

bureau, it would be possible to initiate therapy at a point on the responsibility-balance dimension much farther in the direction of the client than if the therapy takes place in a hospital. Students seek help through counseling but generally do not approach the service with as much dependence as patients in a hospital since they are not only still in the social field of the college campus, but also have not found their own efforts at adjustment so completely futile.

OUTLINE OF RESEARCH PROPOSAL

On the basis of the foregoing discussion, the following research seems appropriate:

1. Identification and initial generalized description of dimensions or continua in the areas of relationship and technique.
2. Development of dimensional scale values (Thurstone method) for therapist responses in successive sections (fourths, fifths or sixths) of transcribed completed series of cases.
3. Analysis of content of patient responses in terms of criteria of change or progress.
4. Graphic representation and correlation of scale-value change for the various continua with quantitative expressions of the criteria of progress.
5. Final determination of the segments of continua of relationship and technique most appropriate with respect to (a) type of case or problem, (b) stage of therapy, and (c) related social fields.

SUMMARY

A basis for integrative or convergent trends among the current disparate emphases in psychotherapy has been described. This basis is found in the proposed identification and quantitative description of continua or dimensions in both technique and therapist-patient relationships. Such continua are suggested as providing a basis for including all the aspects of technique and relationship that have been found effective in psychotherapy. The continuum concept is offered as a potential resolution of dichotomies such as "directive" and "nondirective." The needs of patients should finally be thought of in terms of the appropriate pattern of segments of these dimensions for particular cases, for particular stages of therapy and for variations in the social field within which therapy occurs. Concrete proposals for the application of the continuum concept to research in psychotherapy are offered.

Received August 16, 1949.

REFERENCES

1. AXLINE, VIRGINIA M. *Play therapy.* Boston: Houghton Mifflin, 1947.
2. CURRAN, C. A. *Personality factors in counseling.* New York: Grune & Stratton, 1945.
3. DEUTSCH, F. *Applied psychoanalysis.* New York: Grune & Stratton, 1949.
4. DOLLARD, J. AND MOWRER, O. H. A method of measuring tension in written documents. *J. abnorm. soc. Psychol.,* 1947, **42,** 3-32.
5. EIDELBERG, L. *Take off your mask.* New York: International Universities Press, 1948.
6. MOWRER, O. H. Learning theory and the neurotic paradox. *Amer. J. Orthopsychiat.,* 1948, **18,** 571-610.
7. MURCHISON, C. (Ed.) *Psychologies of 1925.* Worcester, Mass.: Clark Univ. Press, 1928.
8. MURCHISON, C. (Ed). *Psychologies of 1930.* Worcester, Mass.: Clark Univ. Press, 1930.
9. ROGERS, C. R. *Counseling and psychotherapy.* Boston: Houghton Mifflin, 1942.
10. ROGERS, C. R. *Counseling with returned servicemen.* New York: McGraw-Hill, 1946.
11. ROGERS, C. R. The attitude and orientation of the counselor. *J. consult. Psychol.,* 1949, **13,** 82-94.
12. ROSEN, J. N. A method of resolving acute catatonic excitement. *Psychiat. Quart.,* 1946, **20,** 183-198.
13. SEEMAN, J. A study of the process of nondirective therapy. *J. consult. Psychol.,* 1949, **13,** 157-168.
14. SNYDER, W. U. An investigation of the nature of nondirective psychotherapy. *J. gen. Psychol.,* 1945, **33,** 193-223.
15. SNYDER, W. U. *Casebook of non-directive counseling.* Boston: Houghton Mifflin, 1947.
16. THURSTONE, L. L. AND CHAVE, E. J. *The measurement of attitude.* Chicago: Univ. of Chicago Press, 1929.

EGO INVOLVEMENT AND THE PROCESS OF DISENGAGEMENT

IJA N. KORNER
UNIVERSITY OF UTAH

A CURIOUS situation has prevailed in the field of counseling in the last half-century. The client has stood in the focus of interest and, therefore, has been the center of research and discussion. The counselor was accepted as a person who, equipped with proper skills and good personality and judgment, faced the client's problems and attempted to deal with them. It was the psychoanalytical group which insisted that this was a misconception and that the counselor, his personality and problems, were as much a part of therapy as the client.

A survey of the literature of the last ten years reveals significant meagerness of publications dealing with the problem of "the counselor in the therapeutic process."

Neo-orthodoxy in therapy, which is represented by the nondirective approach, emphasizing client-centeredness, repeats the assumption that the therapy situation primarily involves the client. The counselor creates an "objective, friendly, permissive atmosphere." The problem of the counselor's feelings, involvements, needs, and tensions is barely taken into account. In his last article [2], Rogers discusses at length what the counselor should feel and do. He is very taciturn as to how this can be achieved or whether it can be achieved at all. Wyatt [3] takes a very different view on the problem of the counselor. He sees the counselor gratifying needs in the counseling process as well as meeting personal problems. The counselor is as much involved, reactive, and situation-centered as the client. No artificial distinction is made between the client, who needs objective permissiveness, and the counselor, who seems to face little difficulty disseminating these conditions.

This writer agrees, to a considerable extent, with Wyatt's concept of a counselor. Any reflection or interpretation given to the client by the counselor is the result of a selective process. The counselor considers many possible ways of looking at what the client is saying, eliminates a number of them, and finally chooses one particular way of responding. This selection is made by the counselor; it is his choice which determines the nature of the reflection or interpretation. Any reflection or interpretation, whether accepted or rejected by the client, structures the client's way of looking at himself and at others. If he accepts what the counselor has said, this means that he sees himself in the way in which the field has been structured for him. A rejection of the counselor's interpretation or reflection, however, does not necessarily indicate that the client's point of view has not been changed. The mere fact that the client rejects a suggestion coming from the counselor has dynamic implications. It can be assumed that any activity by the counselor structures the client's perceptions and, therefore, to some extent is directive, counselor-involved.

The client has a social and personal relationship with a counselor. He likes or dislikes his counselor according to his needs, and only rarely is neutral to him. The client perceives the counselor as a person, someone whose utterances carry weight—sometimes even the pain of insight. Because of this, the counselor must not perceive himself as an abstractum, thinking in terms of the client's actions and reactions and the significance they might have.

During the interview the counselor is part of a social situation, which, like other social relationships, involves emotional interaction. The counselor affects the client in an individual way, a way which is characteristic of his own personality. The counselor is involved with the client simply because both are human beings obeying the laws of human interaction.

The actions of a counselor with respect to a

client are two-fold: He influences the client consciously by giving reflections and interpretations, and he affects the client unconsciously with the weight of his own personality and personality needs. The nature of the reflections and interpretations is often characteristically influenced by the personality needs of the counselor. The questions then arise: How can the counselor guard himself against the danger of excessively projecting his own personality into the counseling process? What precautions can be taken against counselor involvement?

Some schools of therapy believe that it is imperative for every counselor to undergo a psychoanalysis before engaging in any kind of therapeutic work. Although the writer agrees generally with this assumption, the physical impossibility of translating this idea into reality must also be recognized. In addition, it is questionable whether a psychoanalysis is a prerequisite to therapeutic work done on the counseling level. There is little doubt, however, but what workers dealing with milder adjustment difficulties have to guard themselves against involvement as strenuously as those dealing with complex and severe maladjustments.

It is generally assumed that the recording and replaying of interviews accompanied by discussions with co-workers is one way of guarding against involvement. This kind of analysis makes it possible for the counselor to retrospectively become aware of his often gross projective tendencies. But it is during the time that the interview is actually taking place that the client is exposed to the falsifying factor of involvement. It is during the interview, when the important components of the therapeutic process are in the making, that the counselor needs ways of detecting and dealing with his involvement. Some possible ways in which the counselor might deal with this problem will be considered in the following paragraphs.

During the interview the counselor, apprehensive of his own as well as the client's reactions, should be aware of and familiar with two processes. First, he must recognize indications of unconscious involvement and be aware of their appearances. Secondly, at various times during the interview he must attempt to draw back from the interactive process and determine the nature of his own functioning. It is suggested that this stepping-back process be called "disengagement." It is the experience of this writer that both of these processes can be learned and practiced successfully in the interview situation by any qualified counselor. When handled with proper skill and sensitivity, these processes will not damage or interfere with the counseling interview.

SIGNS OF COUNSELOR INVOLVEMENT

In his interaction with the client, the counselor is often subject to obstructions from within himself. These obstructions are indications of counselor involvement. For example, the counselor finds his thoughts wandering away from the interview. It is difficult for him to concentrate, and he suddenly finds it hard to follow the client. Contrary to his previous experiences with this client, he finds that he cannot completely grasp the material which is being presented. He experiences a lack of sensitivity toward the client. The counselor may think that he is not up to par on this occasion, but it is possible that these manifestations are indications of resistance to some of the material under discussion. Perhaps some specific topic brought up by the client contained material which by association is problematic to the counselor himself. It is possible that the counselor is unaware of his own problem and, therefore, is also unaware of the effects of the client's statements upon him? There is, of course, a possibility that the client's presentation is actually confused. If, however, the difficulty which the counselor experiences in trying to comprehend and follow the client is a sudden one and is followed by disinterest and wandering thoughts, it seems very reasonable to believe that these manifestations are symptomatic of counselor involvement.

Another similar phenomenon is the sudden appearance of unrelated thinking in the counselor's awareness. He may have a sudden, persistent thought which seems to have no connection with the subject under discussion. For example, the counselor may experience an urge to walk out of the room, although there seems to be no obvious reason for this wish. Another instance might be the sudden appearance of images in thinking, the images being vivid, strong, strange, and unrelated to the interview in progress. The counselor may attempt to repress these manifestations, but cannot control their reappearance. These phenomena are di-

rectly related to the counseling process; they indicate the active presence of repressions. The counselor is involved, but he is unaware of his involvement.

At times the counselor may feel that the interview is not moving along as it should. Although the client is not especially reticent or inhibited, he gives little information. During the interview one impasse follows another. The counselor feels at a loss as to what to do next. Although he feels that he should do something, he lacks his usual, calm patience. Perhaps he is not aware of the strong feelings of insecurity he experiences in his relationship with a given client or during one particular interview. The sudden urge to be active in an interview, to direct it, may be an indication that the counselor is not detached.

A counselor sees many clients on a certain day, and to his surprise, many of them seem to show similar problems with similar dynamics. He notices that the material brought forth that day by different clients lends itself to similar interpretations; the adjustment difficulties seem to have a common pattern. This similarity of patterning and interpretation is often more than accidental. A careful review of the interviews with a co-worker often reveals a sensitivity of the counselor toward some specific aspect of the clients' problems rather an coincidental similarity of problems.

While the counselor is trained and often can recognize negative feelings toward a client, he is much less sensitive to his too positive feelings. If the client stimulates strong feeling of liking and sympathy in the counselor, these feelings frequently pass unnoticed. The counselor is not aware of this kind of involvement. The counselor may identify with the client in certain respects or he may be attracted to the client and develop strong emotional ties to him. Oversympathy and marked liking for a client falsify the counseling process as radically as hostility and dislike. Signs of overliking are often manifested in the counselor's behavior by overeagerness, increased verbal participation, an urge to introduce himself into the conversation, a sudden directiveness, and many others.

A very real problem to many a counselor is the one of setting his own limits. Some counselors set their limits of therapeutic activities very low. They hesitate and avoid entering the realm of what they label "deeper problems." This avoidance of therapy under the arbitrary heading "deep problem" is often a sign of the counselor's need to avoid involvement to any degree under all circumstances. Other counselors have no hesitation to tackle any problem the client may bring, disregarding the factor of their own competence and preparedness. This, frequently, is indicative of the counselor's need to get involved. Both instances indicate indirect counselor involvement in the counseling process.

The recognition of "signs of involvement" is of little value to the counselor if he is unable to follow these signs through to an understanding of their meaning during the time that the rapidly changing interview is still in progress.

DISENGAGEMENT

In order to more objectively evaluate the dynamics of the interview in progress, the counselor should repeatedly attempt to ask and answer the following questions: Do I notice any signs of disturbance within myself? Am I functioning normally? Do I feel comfortable? In order to evaluate his own feelings and actions in the light of the movement of the interview, it is necessary for the counselor to step back from the situation involving the client and himself. It is suggested that his step be called "disengagement" and that it be made a part of every counseling interview.

Speed and intensity are two important factors in the interview. Speed is defined here as the rapidity with which the client divulges information or deals with it. Intensity is defined as a measure of the emotional involvement a client experiences while divulging or dealing with material. These factors frequently overlap in an interview. In every interview there are high and low points of interrelated speed and intensity. At a low point, the client seems to be at a standstill. At this time he may repeat material or restructure it. His feelings may be at a low pitch. Such points of low intensity and low speed permit the counselor to disengage himself without damaging his rapport with the client. At such times the counselor can take stock of himself and his feelings. He can review the material and determine the adequacy of the reflections and interpretations given to the client. It is possible for him to ex-

amine his own conduct and watch for signs of involvement. The counselor removes himself from the place of an active participant, but he must remain alert to the client's reactions and sensitive to any sudden change in the situation.

If the results of the self-inventory are negative, it may be possible for the counselor to resume his activities with more awareness of the dynamics of the interview and with increased confidence.

In cases where the counselor discovers signs of involvement, he may not be able to discover the deeper lying sources of his disturbance. Without special and prolonged training, this is frequently extremely difficult. He can, however, take adequate precautions which will prevent damaging projections of these disturbances into the counseling process. The counselor, for example, may decide immediately to take a more reserved and cautious attitude toward the client. He may limit the quality and quantity of interpretations to an absolute and well-controlled minimum. He may desist from any step which would increase the speed and intensity of the interview. Last, but not least, the counselor may be alerted to the necessity to consult with associates in order to determine more clearly the nature, significance and the gravity of the disturbances.

SUMMARY

This writer assumes that the client and the counselor both face a new, unstructured, and often difficult situation in the process of therapy. Both personalities are involved. Both have needs to be gratified, adjustments to make. Research has given valuable information on some aspects of the client's needs, feelings, and reactions. Research now has to investigate the personality and the function of the counselor. The counselor and the client are made of the same material, and both are alike in many more ways than they are different. We have to learn the properties and functions of the tool called "the therapist" in order to understand more fully his effect on the client.

The counselor in the counseling process cannot help but be affected at various times by the client's reactions. It is important that these signs of involvement be recognized in order that projections of the counselor's problems into the counseling process be forestalled. The process of disengagement is helpful in determining counselor involvement. The recognition of involvement enables the counselor to take precautions against projections.

Received July 26, 1949.

REFERENCES

1. ROGERS, C. R. *Counseling and psychotherapy.* Boston: Houghton Mifflin, 1942.
2. ROGERS, C. R. Attitude and orientation of the counselor in client-centered counseling. *J. consult. Psychol.*, 1949, **13**, 82-94.
3. WYATT, F. The self-experience of the psychotherapist. *J. consult. Psychol.*, 1948, **12**, 82-87.

AN INVESTIGATION OF NONDIRECTIVE GROUP THERAPY WITH STUDENTS IN ACADEMIC DIFFICULTY[1]

WILLIAM D. SHELDON
SCHOOL OF EDUCATION
SYRACUSE UNIVERSITY
AND
THEODORE LANDSMAN
DEPARTMENT OF PSYCHOLOGY
UNIVERSITY OF DELAWARE

THE success of nondirective therapy in helping maladjusted college students to move toward better adjustment has directed attention to its possibilities for students with academic difficulties. It is suggested that these students might well need, for improvement in learning, a better understanding of themselves as well as training in the skills of study technique. A study of the educational literature reveals a number of individuals who are investigating the nondirective group therapy approach to education. These studies have contributed much toward the early exploration of this seemingly profitable methodology. Among the recent writers who have presented studies in this area are Shedlin [7], Peres [5], Baruch [1], Faw [2], Schwebel and Asch [6], and Landsman [3].

Arthur J. Shedlin [7] describes a course in which the instructor or counselor "attempted to conduct all sessions in a consistently permissive student-centered manner allowing the initiation and direction of the discussion to be completely determined by the values and interests of the group." The results were reported as favorable with only one negative response on self-evaluation made at the termination of the course. Shedlin [7] felt that the favorable results were due largely to the attitude of the instructor and the climate which he developed.

One of the more detailed studies of nondirective therapy in the group form is the investigation by Peres [5] which has analyzed the content more than the results of this process. Baruch [1] has used similar techniques with classes although she has concentrated most of her study in the therapeutic results with less attention paid to the more general educational values. Faw [2], and also Schwebel and Asch [6] report favorable results in psychology classes in tests of subject matter and in subjective reports of class members.

By contrasting a nondirective counseling class with a conventional mental hygiene class, Landsman reports that students reacted more favorably to the nondirective class with respect to personal growth, information, long term value, and other categories. Landsman's [3] evaluation methods are essentially student-centered, lacking in objective criteria, and are primarily subjective reports of students which are analyzed quantitatively. While this is consistent with the educational methods employed, it departs to an extent from traditional evaluative methods.

BACKGROUND AND PURPOSE OF THE STUDY

For several years classes in Academic Methods have been held for Syracuse University students. The students attending these classes were those who through pretests, or actual failure in academic work, demonstrated a lack of mastery of reading and study skills. The classes were conducted in a conventional manner. The instructors held both lecture and laboratory sessions during which students were taught to read more effectively and to apply well known study skills to their academic work. The continuous search for better classroom

[1]This research was sponsored by the School of Education of Syracuse University and is adapted from a paper presented at the 1949 meeting of the American Psychological Association in Denver, Colo. before Division of Educational Psychology.

methods in these Academic Methods courses prompted this study. This research is aimed at exploring the utility of nondirective group therapy for the Academic Methods classes.

The emphasis of nondirective therapy is in the understanding and acceptance of the self or the self-concept. Some few modern authors, Lecky [4], Snygg and Combs [9], have hypothesized concerning the crucial importance of the student's attitudes toward himself. They indicate that, at the root of many a student's failure is not the inability to master study methods but a concept of self which is inadequate. Snygg and Combs [9] stress the importance of the self concept in all phases of adjustment and specifically with regard to its position in determining the success or failure of a student.

Sheldon [8] has added experimental evidence for this point in his study of forty college students with academic difficulties. He found that the failing students studied ranged in intelligence from normal to very superior but that these forty individuals were severely disturbed emotionally. Sheldon has suggested that an intensive survey of the emotional status of entering college students should be made. A better understanding of himself will, it is hypothesized, enable the failing college student to move toward a more adequate self concept or a better academic adjustment. The resulting change in attitude should enable the student to improve not only in his academic work and social life but should lead to a greater all around adjustment. In order to investigate this thesis, the study was made.

SELECTION OF SUBJECTS

A group of twenty-eight freshmen, whose performance during their first semester of college was below expectation, were invited to participate in the course in Academic Methods. According to the results of their entrance examinations all of the students in the study were potentially capable of succeeding in college work. These students were divided into two classes using the matched pair technique. The students were paired roughly according to intelligence as measured by the Ohio State Psychological Examination, general adjustment by the California Test of Personality, and reading ability with the Iowa Silent Reading Test, and grade point average.

According to Table 1 the variances were homogeneous and there were no significant differences between the means of the two classes for the California Test, the Ohio State Psychological Examination, and the grade point averages prior to the study.

TABLE 1

THE *t* VALUES BETWEEN MEANS AND THE *F* RATIOS BETWEEN VARIANCES FOR TESTS ADMINISTERED BEFORE EXPERIMENT

Test	df	*F*	*t*
California Test of Personality	26	2.133	.6048
Ohio State Psychological Examination	26		
Iowa Silent Reading Test	26	1.295	.0909
Grade Point Averages	26	1.3504	1.3793

METHODS OF CONDUCTING CLASSES

The twenty-eight students met three times a week in Academic Methods. The two groups met together each Monday and were lectured on such subjects as the use of time, comprehension, and rate of reading, memorization, concentration, notetaking, and studying for examinations. The lecture was followed by a question and answer period, demonstrations, experiments, and tests. At this joint weekly meeting the conventional method of instruction, lecture, and discussion was used.

The total group was then divided into two separate groups which met two additional times each week. Group A, the control group, was taught in the conventional lecture and discussion manner by the instructor of the joint sessions. The methods included lectures and directed group discussions by the instructor, required reports made by students, supervised practice in reading and study skills.

Group B, the experimental group, spent its two periods in nondirective group therapy sessions led by a competent nondirective therapist. Recordings were made of both groups throughout the semester using the Sound-Scriber.

THE NONDIRECTIVE GROUP THERAPY SESSIONS

Since there is practically no standardization of meaning for the term "nondirective group therapy," it is advisable to clarify specifically the methods used in the class sessions. Em-

phasis is primarily placed upon an "atmosphere" which is created in the classroom situation. This is both a difficult and limitless task since the atmosphere must be defined in terms of permissiveness perceived by each individual member of the group. The instructor carefully attempted to remove the traditional classroom threats such as the authoritarian view of the instructor, the fear of criticism, and the necessity for meeting arbitrary standards.

Classroom techniques rested largely upon the now traditional "reflection and acceptance of feeling." It is to be emphasized that although even now the authors lean toward a broader and more liberal interpretation of nondirectivity in the classroom, at the time of the study a more rigid concept was adhered to. The direction of the techniques employed was clearly toward increased independence of the individual student. This is distinguished from the control situation in which students relied upon the instructor to provide the best techniques for studying.

The participation of the instructor was primarily in (a) creating the atmosphere of acceptance, permissiveness and warmth, (b) reflecting feelings of students as expressed in the classroom discussion, (c) structuring the course verbally when needed. This includes "putting over" the point that the course would be different from others, the instructor would not lecture, etc., and (d) providing materials and equipment but only when specifically requested to do so by a student.

Members of the nondirective class were not discouraged from dropping out of the course when they wanted to. A classroom situation where a number of students were working out crossword puzzles in the college newspaper was not unusual. At the same time, not unusual were situations where the students were painstakingly discussing and working out problems surrounding negative attitudes toward instructors, feelings of deep inadequacy, fear of examinations, dissatisfaction with high school preparation, resistance to family "pushing," and social and personal adjustments.

Each member of both groups was asked to submit a written report or a reaction report once a week, containing his reaction to the methods used in his group. Individual interviews were given to members of either group upon request. A personal evaluation of the entire course was requested from each student at midsemester and end of the semester.

TREATMENT OF THE DATA AND RESULTS

At the close of the semester, all students in the study who could be contacted were retested with the California Test of Personality and the Iowa Silent Reading Test. Grade point averages, the traditional college evaluating device, were accumulated for the close of the semester during which the study was made. One year follow-up studies were made with respect to whether or not the students survived the academic storms which followed for them.

The weekly reaction reports were analyzed for the purpose of detecting trends in the development of individual students or in each group as a unit. These reports were studied together with the recordings which were made on corresponding dates in order to determine any existing relationships between the method of conducting the sessions and the reaction of the classes.

As indicated in Table 2, variance ratios for the two groups were found to be homogeneous but no significant differences between the means were found after the experiment for either the California Test of Personality or the Iowa Silent Reading Test.

However, variance ratios were homogeneous and a difference significant at the one per cent level of confidence was found for the academic grade point average immediately after the experiment in favor of the experimental or nondirective group.

TABLE 2

VARIANCE RATIOS AND *t* SCORES FOR TESTS AND GRADES FOLLOWING THE EXPERIMENT

Test	df	*F*	*t*
California Test of Personality	14	1.73	.00849
Iowa Silent Reading Test	17	1.6296	1.0833
Grades	20	1.3081	6.5**

**Difference significant at 1 per cent level in favor of the non-directive or experimental group.

Control Group A (Traditional). Seven of the original fifteen students in Group A, the control group, dropped out of college, five because of poor grades and two left of their own accord. These two received poorer grades in the fall term of 1947 than they had received at the close of the previous experimental term

TABLE 3

SUMMARIES OF THE WEEKLY REACTION REPORTS OF FOUR MEMBERS OF THE CONTROL GROUP

Date	STUDENTS A	B	C	D
3/3	The course has been conducted well enough to the present and the exercises have proved helpful.	I'm pleased with the classes and think I shall benefit from them.	As far as I can see not too much has been accomplished.	I find myself progressing slowly but surely. If we had a five day schedule we'd progress more rapidly.
3/7	We study so much for detail at the university that we know little about grasping concepts.			I personally think course is all right. I'm beginning to notice an improvement in my comprehension and reading rate.
3/14	No reaction.	We ought to get started on some of the major topics which are to be covered.	I'm rather well pleased with the course so far. I believe that my reading skill has improved slightly.	So far I see I have a tendency to improve.
3/21	The course is now broadening considerably.	This morning's session was very interesting and helpful.	I find this one of my most interesting and beneficial courses.	At present no criticism.
3/28	No comment.	None.	I haven't progressed as rapidly lately. The discussions are helpful but they don't help my reading speed and comprehension.	More reading tests.
4/23	Would like a discussion of note taking on reading material which contains quantities of factual material.	We should discuss how to prepare for exams.	My rate has gone up 222 with 100% comprehension. I don't think this improvement is very great and I'm still much too slow a reader.	Would like more forced time rapid reading material.

(Spring, 1947). There were eight students in the Control Group still in college at the close of the spring term of 1948. Five showed some improvement in one or both of the semesters following their attendance in the Academic Methods class and three showed a decrease in grade point average.

Experimental Group B (Group Therapy). Three of the twelve dropped out of college.[2] Two dropped out at the close of spring, 1947, session and a third was dropped by the college at the close of the 1948 school year. Of the nine who remained in college for two semesters following the Academic Methods classes, three showed an increase in grade point average, one had the same grade point average at the close of the second semester following the course, and the remaining five showed a decrease.

[2]Original group was 13, but one individual dropped the class immediately, leaving an N of 12 for this group.

TRENDS

An analysis of the reaction reports submitted by the students in the Academic Methods course revealed certain trends. The reaction of only four typical students from both the control and experimental groups is presented in Tables 3 and 4.

Control Group A. The majority of the stu-

TABLE 4
SUMMARIES OF THE WEEKLY REACTION REPORTS OF FOUR MEMBERS OF THE EXPERIMENTAL GROUP

Date	STUDENTS A	B	C	D
2/21	No value to us. Value for instructor.	So far nothing. I'm expecting bigger and better things.	An intensely intriguing varient from contemporary school rituals. Class not adapted to new environment.	Last two sessions a total waste of my time. My marks aren't so good that I can waste a study period.
3/3	Not as much tangible work as I expected. Help received from class discussion but would prefer full hour of work.	I am familiar with what we are discussing. Nothing solid has been given us to stand on.	Method of instruction caused considerable friction in class. Stimulus to make me follow correct study habits is lacking, I feel.	
3/7			My book has taught me 90% of what I know. This course has slight positive value.	Very little progress has been made toward helping the class. The course has great potentialities.
3/14	Class discussion a lot better. Reading test good, would like more.	Recordings were amusing but helped only one person.	Animated discussion in class. The future promises.	Course seems to be developing a little better, but still seems to waste time.
3/21		My roommates tell me I'm studying more, but I don't believe it is the result of this course.	Some light is dawning regarding my confusion in the quest of education.	
3/28	Examinations will help if continued every session.	No comment.	Delightful material for today's comprehension check.	The class seems to be progressing more. If we can have daily tests, I believe everyone will benefit.
4/11	Reading tests show my work is average. Not much improvement, I feel it will take time.	I don't believe my study habits have improved. I'm studying more, but my marks	Excellent relaxation Most interesting item came up very late in class.	The class has come to a standstill. A great deal of repetition and too much silence.

dents in the Control Group seemed highly pleased by the way in which the course was conducted and felt that they were getting something definite from it. They were pleased with the reading tests and the positive direction the course was taking. These attitudes were still maintained at the middle of the semester. Toward the end of the sessions, however, there appeared in the majority of the reaction reports a feeling of dissatisfaction and some expression of insufficient progress.

Experimental Group B. In the experimental group the entire class (with one exception) was initially antagonistic to the method of instruction. They demanded specific information and definite directions and suggested the instructor was lazy and evading all issues. They felt that they were wasting time and that the

views of their classmates could not contribute toward improving their methods of study. A glance at Table 4 will give the reader some idea of how strong some of the reactions were.

About the middle of the semester one finds a slight change in the reactions of the students. Table 4 again indicates the change in feeling. In the entries for March 14 they state: "Class discussion a lot better," "Course seems to be developing a little better," and "Animated discussion in class," etc. Although they had not fully accepted the new method of instruction even at the end of the course, their comments indicated that they felt more kindly about it yet hesitated to admit that it had really been a help.

SUMMARY

A study is presented of twenty-eight students divided into two "Academic Methods" classes. One group is conducted in a primarily nondirective manner while the other group is conducted in the traditional lecture-discussion manner. Prior to the study the groups did not differ significantly with respect to test scores on the Iowa Silent Reading Test, California Test of Personality, Ohio State Psychological Test, and grade point averages. Following the study, the groups did not differ significantly in Iowa Silent Reading Test scores and California Test of Personality scores. The groups did differ significantly with respect to grade point averages at the conclusion of the semester in favor of the nondirective group. One year follow-up studies revealed that forty-seven per cent of the conventional group and twenty-five percent of the nondirective group had left school.

Received August 4, 1949.

REFERENCES

1. BARUCH, DOROTHY W. Incorporation of therapeutic procedures as a part of the educative process. *Amer. J. Orthopsychiat.*, 1942, **12**, 659-665.
2. FAW, V. A psychotherapeutic method of teaching psychology. *Amer. Psychologist*, 1949, **4**, 104-109.
3. LANDSMAN, T. A study of non-directive group psychotherapy as an educational method. Unpublished M.A. Thesis, Syracuse Univ., 1947.
4. LECKY, P. *Self consistency.* New York: Island Press, 1945.
5. PERES, HADASAH. An investigation of nondirective group therapy. *J. consult. Psychol.*, 1947, **11**, 159-172.
6. SCHWEBEL, M. AND ASCH, M. J. Research possibilities in non-directive teaching. *J. educ. Psychol.*, 1948, **39**, 359-369.
7. SHEDLIN, A. J. A student-centered class. *Personal Counselor*, 1947, **2**, 116-131.
8. SHELDON, W. A study of college students with scholastic Difficulties. Unpublished Ph.D. Dissertation, Syracuse Univ., 1949.
9. SNYGG, D. AND COMBS, A. W. *Individual behavior.* New York: Harper, 1949.

THE IMPLICATIONS OF CERTAIN ASSUMPTIONS INVOLVED IN THE USE OF THE THEMATIC APPERCEPTION TEST[1]

J. R. WITTENBORN
YALE UNIVERSITY

WHEN verbalizations produced in response to the TAT cards and other projective tests are used by clinicians as an evidence for important characteristics of the patient, an employment of two assumptions is involved:

I. That superficially similar productions sample traits or relatively fixed classes of the patient's behavior.

II. That certain superficially dissimilar productions are "dynamically" related to each other.

In addition to Assumptions I and II, projective tests require an assumption which is implicit in all testing procedures, i.e. that the behavior elicited by the tests is a sample of a kind of behavior which may to some degree be elicited from the individual by numerous situations, including both the naural social situations in which behavior has a practical import and the artificial testing situations designed for the purpose of eliciting behavior under standard conditions only. Responses elicited by a test may be conceived as relevant to a universe of possible responses elicitable by situations (including the test situation) which have certain common response-eliciting properties.

Assumption I, with its emphasis on sampling trait behavior, is implicit in two different interpretive practices in clinical psychology. The most common interpretive practice emphasizing a sampling of traits is the one wherein the clinician looks for repeated manifestations of a class of behavior to some degree characteristic of individuals in the society from which the patient is drawn. Some of these classes of behavior, such as mental abilities, have been subjected to formal definition and systematic exploration. Others, such as passivity, hostility, and conflict, are important in clinical jargon but ill-defined and unexplored. Perhaps, originally these words were employed only to describe the manner in which individuals reacted to specific situations. At present, however, these words are also commonly used to designate attributes which characterize individuals in a variety of situations and which are considered to be more or less characteristic of all individuals.

Just as regularities in the behavior of groups of individuals have led to the use of attributes as an economical and uniform means of describing them, other regularities which are uniquely characteristic of the behavior of a particular individual are also often ascribed considerable significance by the clinician's intensive study of that individual. This second interpretive practice with its emphasis on the sampling of idiosyncratic behavior often leads the clinician to a consideration of features of the individual's behavior which may comprise the nexus of his problem but which would not be discovered if only broad, culturally relevant traits were recognized.

Assumption II recognizes that psychologists commonly regard certain superficially dissimilar behavioral incidents or manifestations as functionally related to each other and accept them as an evidence for a broadly integrated pattern of behavior. The principles by which these different manifestations are bound together are often referred to as behavioral dynamics. The clinical psychologist who emphasizes his awareness of the interrelationships between symptoms or other superficially different behavioral manifestations is often said to be "dynamic" or "dynamically oriented." This is particularly likely when he presumes to know the direction of influence and speaks of causal

[1]The writer wishes to acknowledge gratefully the clerical assistance of the Institute of Human Relations, the U. S. Public Health Service, and Dr. Malcolm Williams of the U. S. Coast Guard Academy.

factors (e.g. unconscious motivations). The rules or principles which descibe functional or dynamic relationships are for the most part not clearly stated and psychologists differ in the degree to which they employ various principles and in the manner in which they interpret them. As a matter of fact, the most thorough-going and prolifically described dynamic behavior theory, the psychoanalytic theory, is made up largely of words which are ill-defined and concepts which have uncertain implications.

Psychoanalytic theory enables the clinician frequently to predict the kind of behavior which may be expected from a patient and also to specify types of behavior which are improbable, but the manner in which the clinician does this is not rigorously deductive. In many important instances, the rules or principles actually employed are either unanalyzed or unverbalized so that one says, "I have a feeling that" more often than "I deduce from the foregoing facts and principles that. . ." Because of this, attempts to establish the validity of dynamic principles may be more dependant on the intuition of the experimenter who states the hypotheses than on the adequacy of the theory to which the hypotheses purport to be relevant.

For the purposes of the present discussion, it is proposed that three different types of considerations may be employed in rationalizing the practice of basing general statements concerning the individual's adjustment on the manner in which he responds to the TAT cards. Specifically, (1) responses may be ascribed a practical significance if they are manifestations of a behavioral attribute of general interest; (2) if they occur with such frequency in the productions of the individual as to suggest that they are uniquely characteristic of him; or (3) if they have a status with respect to some system of behavioral dynamics. Considerations (1) and (2) employ Assumption I whereas consideration (3) is based upon Assumption II. If the first type of consideration is validly employed and traits which may be used to distinguish between individuals in a group are directly manifested in responses to the test cards, it would be predicted that TAT responses which express some of the most common traits or attributes would occur with a discernible prevalence among the productions of individuals. It would be predicted, moreover, that the manifestation of a given prevalent form of response to one card would be correlated with a manifestation of that form of response occurring in other cards. If such a consistency of response tendency could not be discerned, one would be inclined to suspect that there was no trait-like consistency in the behavior of individuals with respect to the types of responses sampled by the different cards.

The second type of consideration which may be employed in rationalizing the use of the TAT would not contribute to the prevalence of any form or forms of response. Moreover, the personally unique or idiosyncratic nature of the response would not contribute to intercorrelations among different manifestations of the type of response in question, i.e. a particular manifestation of the type of response in question would not characterize more than one or two members of most samples, and as a consequence this type of response would not provide a basis for pairing the individuals; accordingly, no correlation between various manifestations of a given mode of response would be possible. Although the second type of consideration is not a subject of the present analyses, it is feasible to test the validity of this type of consideration and such tests will provide the subject for a future report.

If the development of and interrelations among behavioral manifestations are controlled by or describable in terms of a set of laws or principles of behavior which specify the relationship between the individual's behavior and the conditions of his existence, it would be expected that similarities in the environments of a sample of individuals would produce similarities between individuals with respect to their behavioral manifestations, including those elicited by projective tests. Presumably, therefore, the third consideration would permit a classification of the responses of a group of individuals. The lawful nature of behavior is emphasized in the third consideration because the types of behavior which are lawfully interrelated may actually be superficially dissimilar in their appearance and their interrelationships may not be congruent with any ordinary conception of behavioral traits or attributes. Although the trait concept may be most readily applied to the similarities

involved in the first and second considerations, it is not implied that the types of behavior involved in the first and second considerations are lawless or unpredictable in their development.

Although a direct validation of the clinical use of the TAT may be readily conceived, such direct validating procedures are difficult to execute because they require the development of criteria, a task which is ordinarily so lengthy and expensive as to make desirable preliminary studies which indirectly explore the probable validity or invalidity of plausible assumptions. The purpose of the present report is to provide an evaluation of two hypotheses which are generated primarily by Assumptions I and II, respectively:

Hypothesis A: Responses to the TAT which are similar to each other with respect to class of behavior involved will be significantly related to each other more frequently than the responses among an unselected group are related to each other.

Hypothesis B: Responses for which a functional or dynamically determined complementary response may be specified will be significantly related with their complementary reresponses more frequently than unselected responses will be related to each other.

Inasmuch as the interpretation of Hypotheses A and B is not a completely objective matter, it is possible that the outcome of an analysis such as the present one is influenced by the responses selected by the investigator. The response selections themselves may be described fairly well, however, and it seems probable that the general procedures employed may be made sufficiently clear to permit an effective reproduction of the conditions of the present analysis.

It is emphasized that it is not the purpose of the present study to test any particular dynamic theory or its implications. It is intended to examine only certain restricted consequences of Assumptions I and II. These assumptions are considered by the writer to be implicit in the ordinary use and interpretation of the TAT, projective tests in general, and other clinical material. The method of analysis employed in the present study could, however, be adapted to the purpose of testing specific interpretations generated by psychodynamic theories of human behavior. As a matter of fact, in testing Hypothesis B, the specific response groups chosen for the statistical tests will have a greater relevance to some dynamic principles than to others. No serious attempt will be made, however, to identify the theoretical status of any of the response groups.

SUBJECTS AND PROCEDURES

The subjects who provided the necessary data for the present analysis were a group of 100 Yale undergraduates who consulted the writer voluntarily. The nature of the problems presented and the therapeutic procedures employed varied widely. At least one-half of the students could be described as chronically neurotic, but only a very small fraction of them had at any time experienced acute episodes or presented psychotic features.

It is quite possible that the sample of subjects providing the data may determine the relative status of Hypotheses A and B. For example, it is possible that the so-called trait behavior with its stereotyped manifestations becomes more conspicuously characteristic of patients as their illness becomes more severe. As a matter of fact, it does not sound unreasonable to say that for the most part, broadly generalized patterns of reaction characterize the behavior of psychotics, and a lack of discriminating behavior is an outstanding characteristic of severe psychopathology. In most, but certainly not all cases, it also seems characteristic that the severely affected patient's behavior is overgeneralized in such a manner as to obscure the functional interrelationships between his different behavioral manifestations and, as a consequence, a verifiable dynamic formulation of his behavior is often tediously protracted if not infeasible. Among patients such as those comprising the present sample, who for the most part are moderately neurotic or are troubled with transient neurotic symptoms, the types of superficial similarities in behavior which lead to broad descriptive classifications or trait designations are not arrestingly conspicuous, and functional interrelationships among many superficially different behavioral manifestations seem to obtrude themselves on the therapist in a manner which is less obvious among most psychotic patients. Although the foregoing pattern of speculation suggests that the present sample favors the verification of Hypothesis B more than Hypothesis A, this

is certainly not a matter of fact and may not be verifiable. Nevertheless, it is urged that the results of the present analysis be generalized only to nonpsychotic samples of intelligent young men who for the most part are free from severe psychopathology.

The response categories employed in a study such as the present one may be expected to influence greatly the outcome of the analytical procedures. Inasmuch as the purpose of the investigation was to explore the implications involved in the clinical use of the TAT, it was decided to employ categories which for the most part were clinically relevant to problems presented by the subjects. Due to the nature of the statistical analyses to be employed, it was decided to use only the forms of responses most prevalently provided by the sample of subjects. Inasmuch as the manifestations of different forms of response were to be intercorrelated with each other, it was imperative that a form of response employed be relatively common, easy to define, readily recognized, and not likely to be confused with some other response form. These limitations exclude many clinically useful and theoretically important forms of response from the present analysis.

Several systems for classifying TAT responses are described in the literature. For the most part, however, these groupings do not enjoy a widespread clinical use and, as a matter of fact, many of them comprise such particularized aspects of the patient's responses that the clinically significant characteristics of the re-

TABLE 1
A Classification of the Response of One Hundred Students to Certain TAT Cards

Card	Figures	Role Ascribed in Response	Response Group Number	Frequency
7 BM	Boy	Impetuous	1	18
		Compliant	2	28
		Conflict	3	18
	Man	Objective and sympathetic	4	60
7 GF	Girl	Accepting guidance	5	35
		Resistant	6	27
	Mother	Pressing	7	20
		Tolerant	8	51
4	Man	Hostile fighter	9	36
		Conflict	10	26
	Woman	Evil	11	26
		Good	12	67
6 GF	Man	Friendly	13	33
		Suspicious	14	22
		Preying	15	34
	Woman	Evil	16	24
		Good	17	66
6 BM	Mother	Sad, shocked, unrealistic	18	63
	Man	Bearing sad news	19	45
		Confessing guilt	20	17
		Emancipation	21	26
2		Eternal triangle	22	20
		Conflict	23	20
		Peaceful Constructive	24	21
9 BM		Workers	25	49
		Idlers	26	35
13 MF	Woman	Good	27	40
		Temptress	28	19
		Innocent	29	26
	Man	Guilty over violence	30	24
		Guilty over neglect	31	29
		Sex Conflict	32	18

sponses are fragmented. Accordingly, the attractive possibility of using a generally known and standard response classification was foregone for the sake of employing a response classification which seemed clinically relevant to the problems of the patients who provided the data. The response categories here employed were selected from a group mentioned in an earlier publication [2]. Frequency of appearance, not relevance to the hypotheses here presented, was the primary basis of the selection. The response groupings were not based on theoretical considerations but comprise only a convenient grouping of prevalent, clinically relevant response themes.

The foregoing factors led to an employment of the cards and response themes which comprise the list shown in Table 1. The productions were secured by a self-administering TAT procedure which has been described elsewhere [2]. Some of the response classifications given in Table 1 require explanation. For example, for Card 7GF the mother described as pressing (response number 7) may be demanding love from the child, protesting her own love and self-sacrifice, demanding

TABLE 2

AN INDICATION OF THE RESPONSE PAIRS WHICH ARE RELATED AT THE 10 PER CENT LEVEL OF CONFIDENCE (SIGN OF RELATIONSHIP INDICATED IN PARENTHESIS)

Response Group Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1																
2	/															
3	/	/														
4	(–)05	10	—													
5	—	—	—	(–)10												
6	10	—	—	(–)01	/											
7	—	—	—	—	—	01										
8	(–)05	—	—	—	01	(–)01	/									
9	—	—	—	—	—	10	—	—								
10	05	—	—	—	—	—	—	—	/							
11	—	—	—	—	—	—	(–)10	—	—	—						
12	—	—	(–)05	—	—	—	05	—	—	—	/					
13	—	—	—	10	—	—	—	—	—	—	—	—				
14	—	—	—	(–)10	—	—	—	—	—	—	—	—	/			
15	05	(–)10	—	—	(–)10	—	—	—	—	—	—	—	/	/		
16	—	—	—	—	—	—	—	10	—	—	—	(–)05	(–)05	01	—	
17	—	—	—	—	—	—	—	—	—	10	—	—	01	(–)01	—	/
18	—	—	—	05	—	—	—	—	—	—	—	10	—	(–)10	10	—
19	—	—	—	—	—	(–)05	—	—	—	—	(–)10	10	—	—	—	10
20	—	—	—	(–)10	—	—	—	—	—	—	—	—	—	—	—	—
21	10	(–)05	—	10	—	—	—	—	—	—	—	—	—	—	—	—
22	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
23	—	—	05	—	10	—	—	05	—	10	—	10	—	—	—	—
24	—	—	—	—	—	—	10	—	—	—	—	10	—	—	—	—
25	—	—	(–)10	—	—	—	(–)10	—	—	10	—	10	—	10	(–)05	10
26	—	(–)10	—	—	—	—	05	(–)05	—	—	10	—	—	(–)10	10	(–)10
27	—	—	—	—	—	(–)10	—	—	—	—	—	—	—	—	(–)10	—
28	—	10	—	—	—	10	—	—	—	—	—	—	—	—	—	—
29	—	(–)10	—	(–)05	—	—	—	—	(–)10	—	—	—	—	—	—	—
30	—	—	—	—	—	—	—	(–)05	—	—	—	—	—	—	—	—
31	—	—	—	—	—	—	(–)05	10	10	—	—	—	—	—	(–)10	—
32	—	—	—	—	—	—	—	—	—	—	—	—	—	—	01	—
Response Group Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16

TABLE 2 (CONTINUED)

Response Group Number	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
1																
2																
3																
4																
5																
6																
7																
8																
9																
10																
11																
12																
13																
14																
15																
16																
17																
18	—															
19	—	—														
20	—	—	/													
21	(–)10	—	/	/												
22	—	—	—	—	—											
23	—	—	—	—	—	/										
24	—	—	(–)10	05	(–)10	/	/									
25	(–)10	—	10	—	(–)10	—	—	—								
26	10	—	—	—	—	—	—	—	/							
27	—	—	10	—	(–)10	—	—	—	—							
28	—	—	(–)05	10	—	—	—	—	—	/						
29	—	—	05	(–)10	—	—	01	—	—	10	/					
30	(–)10	—	—	—	—	—	—	—	(–)05	05	(–)10	—	10			
31	—	—	—	—	(–)10	—	—	—	10	(–)10	01	(–)01	(–)05	/		
32	—	—	—	—	—	—	—	—	—	—	(–)05	—	10	/	/	
Response Group Number	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32

achievement, or in some uninvited way attempting to structure the child's reaction. The man who is seen as a hostile fighter (number 9) in response to Card 4 may be about to fight or wishing to fight but deterred by the woman or other features of the situation. For this card, the man in conflict (number 10) is uncertain of what he wishes to do; usually, the conflict is concerned with his reaction toward the woman. The good woman (number 12) in response to Card 4 like the good women (response numbers 17 and 27) in response to cards 6GF and 13MF is loyal, sexually moral, and not destructive in her behavior or attitudes. The preying man (number 15) seen in response to Card 6GF may be a blackmailer, a sexual adventurer, or distressing the woman without justification. The suspicious man (number 14) may be a detective, a cuckold, or a niggardly, invidious or irritable individual. The conflict response category (number 23) for Card 2 invariably refers to conflict, uncertainty, or ambivalence on the part of the woman in the foreground. The worker category (response number 25) for Card 9BM refers to themes of any kind wherein the reclining men are seen to have just participated in or to be deservedly resting momentarily from some constructive, nonpunitive task. The idler theme (number 26) is concerned with stories where the men have no real purpose or their purpose is to shirk their responsibilities or obligations for work. Neither of these categories includes fugitive or pursuers. The sex conflict of the man seen in Card 13MF may be

moral regret for a sexual act or uncertainty in the presence of sexual temptation. This response does not include repugnance growing out of the physical features of the situation or any violent or destructive behavior which occurred in the situation.

The statistical significance of the relationship between each of the response themes and every other one was determined by means of the χ^2 test for independence. The significance of all of the important relationships is shown in Table 2. There are 491 possible interrelationships among the 32 response themes listed in Table 1. The alternative roles or themes for any given character are mutually exclusive. Accordingly, the 24 possible interrelationships among mutually exclusive responses or themes are eliminated from consideration so that only 467 of the interrelationships are computed in the present analysis.

Both Hypotheses A and B require that the number of significant relationships among response categories designated as specifically relevant to a particular hypothesis be compared with the number of significant relationships among "unselected" categories. For the purpose of making this comparison, relationships which are "unselected" are defined[2] as all of the relationships between themes or response categories from different cards. There are 438 such relationships given in Table 2, and 21 per cent of these relationships between so-called unselected response or theme groups are significant at the 10 per cent level of confidence. This is twice the number of significant relationships that would be predicted on the basis of chance alone. The fact that the various response categories or themes are significantly related with each other may incidentally be taken as evidence that the various themes may be reliably identified. If no significant relationships had been found, the data would have been meaningless until systematic independent tests of the reliability of the response classifications had been conducted. In general, a lack of significant relationships under such circumstances could be due either to a possible intrinsic lack of relationship or to the possible unreliability of the response classifications. Fortunately, in this study the presence of a goodly number of significant relationships makes a systematic evaluation of the reliability of the response classification unnecessary.

[2]Other means of defining unselected responses are possible, but no reasonable alternative would have changed appreciably the results of these analyses.

RESULTS

Three separate tests of Hypothesis A were attempted. For cards 4, 6GF, and 13MF a pattern of response which identifies the woman as "good" is commonly given. Patients who see the woman as playing a good role in all these three situations are frequently described by clinicians as inclined to regard women as good or to interpret the results in some way equally indicative of a tendency to regard each of these three responses as behaviorally similar. Actually, none of these responses is significantly related with either of the others. It is obvious, therefore, that Hypothesis A does not apply to the "good woman" response category, i.e. the meaning of "good woman" is not the same when elicited by different cards and any interpretive practice which assumes their similarity is unsupported by these data.[3]

Clinicians often speak of conflict as if the word referred to an important, broadly general behavior characteristic of the individual. At times some clinicians are more inclined to speak of the "severity of the conflict" or the prevalence of "evidence for conflict" than to specify the circumstance which elicits conflicting reactions or to emphasize the nature of the conflicting reactions. Conflict is commonly and clearly ascribed by patients to the heroes for Cards 2, 4, 7BM, and 13MF (response group numbers 3, 10, 23, and 32). Nevertheless, only two of the six possible interrelationships among these conflict themes are significantly related at the 10 per cent level of confidence. Although this fraction of significant interrelationships may be slightly greater than the fraction of significant interrelations found among the unselected group of responses, there seems to be little justification for regarding these data as evidence for the appropriateness of Hypothesis A with respect to the TAT responses classified as conflict responses. Whether a state of conflict is as highly generalized as many clinicians presume cannot be answered from these data, but the data do not support an interpretation of the present conflict response themes as being behaviorally similar or as having a common implication for the adjustment of the patient.

Clinicians commonly examine test protocols for evidence of hostility and if they find one or two apparent instances of hostile actions or motives in the projective test protocols, they often speak of mild or moderate evidences of hostility and state

[3]These responses do show significant relationships with other responses, however.

or imply that the patient entertains mild hostile feelings. If, however, reference to hostile or destructive behavior or attitudes occurs frequently in a test record, they may say that the record is full of hostility and claim that the patient is a very hostile person. In such instances, it is evident that the same kind of interpretation is given to a variety of responses which have hostile, rebellious, unfriendly, or in some other respects destructive associations. In the present series of responses, there are several which would often be ascribed a hostile significance. Among them would be the impetuous boy seen in Card 7BM, the resistant girl seen in Card 7GF, the hostile man seen in Card 4, the preying man seen in Card 6GF, the guilt-laden man seen in Card 6BM, and the man guilty over violence seen in Card 13MF. Nevertheless, only two of the 15 interrelations among these six response classes (numbers 1, 6, 9, 15, 20, 30) are significant at the 10 per cent level of confidence. Thus, significant interrelations among response themes which presumably are projections of hostile feelings are no more prevalent than they are among unselected themes. This negative finding certainly cannot be construed as support for Hypothesis A.

Although the results of these three simple analyses cannot be taken as an indication that evidence for Hypothesis A cannot be found among data such as these, they certainly do require the observation of considerable caution in ascribing a common meaning to superficial similarities among responses to different TAT cards.

In organizing specific analyses appropriate to Hypothesis B, groups of responses were selected by means of an informal employment of the following guiding principles:

a. Their correlated appearance should be acceptable as a plausible partial expression of some dynamic principle or tendency.

b. Their correlated appearance should also be acceptable as an expression of a pattern of motivation, conscious or unconscious, frequently observed in samples of patients such as the present one.

For example, a young man whose behavior is impetuous or rebellious in the presence of an older man is not likely to see such older men as objective or sympathetic. On the contrary, it may be predicted that he would tend to regard older men as motivated by a desire to punish, exploit, or in some manner prey upon relatively vulnerable individuals such as women or young men. The pattern of relationships among responses 1, 4, and 15 conforms with this prediction. An inverse employment of the same considerations would result in a prediction that young men who are compliant in the presence of older men tend also to regard older men as objective and sympathetic and would not be inclined to regard men as preying on or exploiting women. The relationships among responses 2, 4, and 15 confirm this prediction.

The foregoing example may be accepted by some readers as a plausible interpretation of certain features of the Oedipal situation. Another feature of this situation leads to a second pattern of prediction. In response to Card 13MF with its varied sexual implications, 18 of the students ascribed a severe ethical sexual conflict to the man. It would be predicted in general that young men who are strongly tempted, but nevertheless guilt-laden and uncertain, in a sexually stimulating situation, would be inclined not to recognize the sexual motivations of the woman but to regard her as innocent. Furthermore, it would be predicted that young men who are uncertain of their own sexual rights and unable to acknowledge the sexual requirements of women would tend to see older men as preying on women and not as objective and sympathetic. The interrelationships among responses 32, 29, 4, and 15 confirm this prediction.

In response to Card 6BM, 26 of the students saw the young man emancipating from the home and the domination of the mother. Whenever the drive toward emancipation and independence is sufficiently strong in youths, they are observed to be more rebellious than compliant, and the home is not regarded by them as a tranquil place where all members are satisfied working toward a common goal. The pattern of intercorrelations (among response numbers 1, 2, 21, and 24) is congruent with this speculation.

The three analyses relevant to Hypothesis B tend to conform with the predictions, whereas the analyses relevant to Hypothesis A offer no support for that hypothesis. As already emphasized, however, the results of the analyses are dependent upon the relationships examined and the reader may scrutinize Table 2 in order to satisfy himself concerning the relative status of Hypotheses A and B. An element of dynamic plausibility may be "read into" a number of the interrelationships. Although it would be extremely hazardous to ascribe any general significance to such inferences, they could, of course, be stated formally and another sample of data gathered and analyzed for the purpose of examining the validity of any inferences that the present data suggest.

Although the present analyses suggest that TAT interpretations based upon a trait concept of behavior should be employed with con-

siderable caution, the possibility of a profitable use of trait concepts is by no means excluded by these limited data and restricted analyses. The analyses do strongly suggest that the most consistent interrelationships among TAT material provided by students can best be understood through an employment of behavioral dynamics. The manner in which the present data challenge the descriptive trait approach to the study of behavior and direct our attention to the importance of the motivations served by the behavior is reminiscent of some of the classical studies of Hartshorne and May [1].

The present analyses provide an interesting methodological exploration. Too many psychologists have resolutely claimed that dynamic material cannot be analyzed by statistical means and that such quantitative statistical analyses are necessarily static and invariably antithetical to a dynamic conceptualization of behavior. It may indeed prove possible to conduct systematic, quantitative explorations of important dynamic interrelationships found in clinical material through an employment of very simple statistical procedures.

SUMMARY AND CONCLUSIONS

The present study comprises a statistical evaluation of data relevant to the following hypotheses:

Hypothesis A: Responses (to the TAT cards) which are similar to each other with respect to class of behavior involved will be significantly related to each other more frequently than the responses among an unselected group will be related to each other.

Hypothesis B: Responses for which a functional or dynamically determined complementary response may be specified will be significantly related with their complementary responses more frequently than unselected responses will be related to each other.

The data comprise a categorization of TAT productions provided by 100 Yale undergraduates in response to 8 selected cards. For the most part the response categorizations were based upon the roles commonly ascribed by the students to the figures appearing in each card. For this reason relatively few response categories were common to two or more cards.

The results of the analyses may be summarized:

1. As a means of evaluating Hypothesis A, three different sets of intercorrelations among superficially similar response categories were examined. In no one of these sets was there a discernible tendency for superficially similar response categories to be consistently interrelated. The failure of the present data to provide evidence in support of Hypothesis A suggests that clinicians should be cautious in ascribing a trait significance to the multiple appearance of TAT response roles or themes which are superficially similar to each other. The relative paucity of significant interrelationships within the sets of response categories relevant to Hypothesis A does not preclude the possibility that some other superficially similar response categories would be positively interrelated and could thereby provide support for a trait emphasis in the clinical use of the TAT productions. The failure of the present data to provide evidence for Hypothesis A may mean, however, that the psychological significance of a given response is contingent upon the context in which the response is elicited and that superficially similar responses to different TAT cards would not be behaviorally similar unless the cards in question were highly similar in their general response eliciting properties, i. e. had in general a similar total meaning for a large number of subjects.

2. Three sets of intercorrelations were examined as a means of evaluating Hypothesis B. The response categories comprising each of these sets were superficially dissimilar to each other, but were considered to be related with each other in a manner consistent with a dynamic interpretation of behavior, e.g. the response categories within a set could be conceived as a partial expression of a single pattern of motivation, conscious or unconscious. The response categories within a set were intercorrelated in a fashion consistent with the requirements of Hypothesis B. There are several specific respects in which this is an important finding:

a. It suggests that the responses elicited by projective test devices may be lawfully interrelated in a manner consistent with a theory of personality.

b. It suggests that clinicians may profit from a self-conscious and deliberate use of per-

sonality theory in interpreting projective test responses.

c. It suggests that the intercorrelations among different categories of responses to projective test material may be used as a method for testing the relative validity of alternative hypotheses in a personality theory.

It is important that the results of a study such as the present one not be overgeneralized and not be incautiously applied to untested situations. The relative status of Hypotheses A and B may vary with the patient group, with the manner in which responses are categorized, and with the relative pertinence to Hypothesis A or B that a given investigator ascribes to a particular response category.

Received August 12, 1949.

REFERENCES

1. HARTSHORNE, H. AND MAY, M. A. *Studies in deceit.* New York: Macmillan, 1928.
2. WITTENBORN, J. R. Some Thematic Apperception Test norms and a note on the use of the test cards in the guidance of college students. *J. clin. Psychol.*, 1949, 5, 157-161.

OBJECTIVE RORSCHACH RESPONSE PATTERNS DIFFERENTIATING ANXIETY REACTIONS FROM OBSESSIVE-COMPULSIVE REACTIONS

SOLIS L. KATES[1]
MICHIGAN STATE COLLEGE

THE purpose of this investigation was to differentiate between the Rorschach test responses of anxiety reaction patients and those of obsessive-compulsive patients who were seeking assistance at a mental hygiene clinic. It was also intended to examine the significance and meaning of certain Rorschach response categories in accordance with their capacity to discriminate between anxiety reactions and obsessive-compulsive reactions.

The patient with an anxiety reaction disorder is suffering primarily from many unrelieved tensions that contribute to his restlessness and that interfere with adequate satisfaction from his achievements. His anxiety is diffuse, is not restricted to specific situations, and is not bound by any psychological defense mechanisms [10]. Both psychological and physiological aspects of anxiety are experienced by the patient. The anxiety reaction patient has defenses that are only partially adequate against his fear of catastrophic breakdown and disaster to his personality structure. These defenses are not sufficient to control and suppress those fears even temporarily [9]. He is also egocentric in that he concerns himself, first of all, with his own affairs before his attention can be devoted to any tasks at hand [7].

The patient with an obsessive-compulsive reaction is characterized by uncontrollable and irrepressible tendencies to say, do or think something in a particular way [1]. These tendencies are said to have value as distractions and as tension-reducing mechanisms which may allay in some manner his anxiety [6]. It has been said that the patient is attempting to satisfy his need for safety and security through these obsessive-compulsive mechanisms [4]. The irrepressible tendency is reported to be more tolerable and less painful to the patient than the unpleasant memories or acts it is reputed to replace [8]. By this process, it is believed overt anxiety is moderately well avoided [9]. The anxiety is apparently under control as a result of the irrepressible tendencies acting as temporary shields against threats to the personality.

It would appear that the degree of inner distress experienced by the anxiety reaction patient is greater than that of the obsessive-compulsive patient and that this difference in degree of inner distress would be manifested on a test such as the Rorschach. Finally, certain Rorschach responses would be associated with the specific adjustment patterns of the anxiety reaction patient in contrast with those Rorschach responses that would be related to the particular modes of adjustment of the obsessive-compulsive patient.

Subjects. The subjects of this investigation were patients at a mental hygiene clinic of the Veterans Administration. Twenty-five patients exhibiting symptoms of an anxiety reaction and twenty-five patients with symptoms of an obsessive-compulsive disorder were selected and agreed upon by three psychologists as representative of these reactions. They were chosen on the basis of symptoms alone, independent of test results and previous diagnostic classifications. There was no significant difference between the mean IQ and the mean chronological age of the two groups.

Method. The Rorschach test was administered to the subjects and their Rorschach re-

[1]Miss C. Mildred Ceres, clinical psychologist, and Mr. Melvin Kornreich, clinical psychology trainee, Brooklyn VA Mental Hygiene Clinic, provided considerable assistance in planning and carrying through this investigation. Interpretations and conclusions do not reflect opinions of the Veterans Administration.

sponses were evaluated by means of the Munroe Inspection Technique [5]. The Munroe Inspection Check List consists of 28 categories embodying the major Rorschach scoring symbols. Specific numbers of entries were made for each response category according to the extent of its deviation from predetermined clinically established normal limits. If within these normal limits, no entry was made. If outside these normal limits, one, two, or three entries were made, dependent upon the extent of the deviation. The total score was the sum of the number of entries. A high score signified many signs of maladjustment and much inner tension as reflected in the Rorschach responses while a low score indicated few signs of Rorschach maladjustment and little inner tension.

Results. The mean number of signs of Rorschach maladjustment as indicated by the Munroe Inspection Technique was 10.0 for the obsessive-compulsive patient, with a standard deviation of 3.6. For the anxiety reaction patient, the mean Munroe Inspection score was 13.2 with a standard deviation of 4.0. The difference between these mean scores was significant at the one per cent level.

The mean number of the responses given to the Rorschach cards by the obsessive-compulsive patient was 37.6 and the standard deviation was 18.9. The anxiety reaction patient's mean number of responses was 18.8, standard deviation of 8.5. There was a significant difference at the one per cent level between the means.

Table 1 lists six Rorschach response categories that differentiated significantly between the obsessive-compulsive and the anxiety reaction patient. The remaining response categories of the Munroe Inspection Check List were nondiscriminating.

Interpretation. The greater number of Rorschach signs of maladjustment of the anxiety reaction patient may be attributed to his more persistent feelings of being under attack, of being helpless and worthless. The loss of the affection and esteem of others is gravely feared. The anxiety reaction patient does not have the defense mechanisms of the obsessive-compulsive patient that are vigorous reactions against similar threatening and helpless feelings. These unacceptable feelings are counteracted by the use of irrepressible tendencies which help reduce inner distress in the obsessive-compulsive patient. This finding should not be construed that, given an equivalent degree of anxiety, the obsessive-compulsive patient, by utilizing

TABLE 1

DISCRIMINATIVE RORSCHACH RESPONSE CATEGORIES ON THE MUNROE INSPECTION CHECK LIST

Rorschach Response Category	Obsessive Compulsive Patients N = 25	Anxiety Reaction Patients N = 25	Chi-Square	Degree of Freedom	Level of Significance
Rejection of one or more Rorschach Cards					
Rejection	3	10	4.9	1	5%
Non-Rejection	22	15			
Rare Detail (dr) Responses					
Less than 10%	11	19	5.6	1	2%
10% or more	14	6			
Exclusive Form (F) Responses					
Less than 50%	19	10	6.2	1	2%
50% or more	6	15			
Inanimate Movements (m) Responses					
Less than two	18	24	4.2	1	5%
Two or more	7	1			
Definite Form-Color (FC) Responses					
More than one	15	7	5.2	1	5%
One or less	10	18			
Color (FC, CF, C) Responses					
More than two	14	7	4.02	1	5%
Two or less	11	18			

his defense mechanisms, will suffer less inner distress than the anxiety reaction patient.

The greater number of responses of the obsessive-compulsive patient, in contrast to the anxiety reaction patient, may be explained by his greater need to demonstrate his adequacy. On the other hand, the anxiety reaction patient is so engrossed in himself and in his efforts to avert a catastrophic breakdown that he cannot devote himself to the task at hand. The task at hand is not the center of interest; rather the necessity to arm himself against imminent breakdown is predominant.

The anxiety reaction patient, by rejecting more Rorschach cards than the obsessive-compulsive patient, indicated that he adjusts to many situations by a process of rejection and withdrawal. On the other hand, the obsessive-compulsive patient may be said to adjust to many situations by a drive for completeness and quantity.

The greater tendency of the obsessive-compulsive patient to give rare detail (*dr*) responses, when contrasted with the anxiety reaction patient, reflected his meticulousness. There seemed to be a drive within the obsessive-compulsive patient to arrange his world in an orderly fashion by taking into consideration the many minute details of his life. Furthermore, the obsessive-compulsive patient probably feels that social value accrues by a mental approach focusing on minute details. Goldfarb's investigation discovered a similar trend in the utilization of the rare detail (*dr*) response by the obsessive-compulsive adolescent [2].

The exclusive form (*F*) responses have been said to demonstrate the degree to which the individual was prone to control his more personal or spontaneous reactions [3]. It would, therefore, appear that the anxiety reaction patient, more than the obsessive-compulsive patient, fears his spontaneity and therefore seeks refuge in the more formal aspects of behavior. Not having sufficient defenses for the control of the fears aroused by his spontaneous reactions, the anxiety reaction patient resorts to an impersonal, matter-of-fact way of dealing with situations. The obsessive-compulsive patient, too, is fearful of spontaneity but not to the extent of the anxiety reaction patient. For he has erected some defenses that cope, even if temporarily, with the anxiety provoked by spontaneity.

It may not, therefore, be true that obsessive-compulsive elements are present whenever the percentage of exclusive form (*F*) responses is fifty or more as Klopfer suggests [3]. The use of fifty per cent or more exclusive form (*F*) responses in the Rorschach test should perhaps be more generally interpreted as a need to be impersonal and matter-of-fact to avoid anxiety and possible acute breakdown.

The inanimate movement (*m*) responses may be interpreted, as Klopfer suggests [3], as the experiencing by an individual of uncontrollable inner forces working upon him. Their more frequent presence in the responses of the obsessive-compulsive than in those of the anxiety reaction patient would support this interpretation. The inanimate movement (*m*) responses may be construed then as being indicative of obsessive-compulsive trends. It also appears that much inner tension may be present without the appearance of inanimate movement (*m*) responses.

The obsessive-compulsive patient's capacity for emotional adjustment to outer reality is greater than that of the anxiety reaction patient as shown in his greater use of form-color (*FC*) responses. The anxiety reaction patient fails to display much socially acceptable affect because he is so greatly concerned with himself and the anticipation of threat from every quarter. Furthermore, he is so fearful that emotional relations with others may result in additional anxiety that he limits the extent of his warm, spontaneous contacts. The same reason would apparently explain the lack of use of an adequate number of color (*FC, CF, C*) responses by the anxiety reaction patient when contrasted with the obsessive-compulsive patient.

Summary. It was pointed out that the anxiety reaction patient had more persistent feelings of helplessness and worthlessness, which were manifested in a higher Rorschach maladjustment score. The obsessive-compulsive patient had more defense mechanisms which helped to reduce his feelings of inner distress, probably resulting in a lower Rorschach maladjustment score. The much larger number of Rorschach responses of the obsessive-compulsive patient, when contrasted with the anxiety reaction patient, may be due to his greater need

to demonstrate his adequacy and to offset feeling of inferiority. It may also be due to the inability of the anxiety reaction patient to get away from his own problems. Similar reasons were apparently associated with the greater number of rejections of Rorschach cards by the anxiety reaction patient when contrasted with the obsessive-compulsive patient.

The obsessive-compulsive patient demonstrated a greater tendency to consider and mull over the minute and relatively unimportant details of life. The anxiety reaction patient was more inclined to utilize impersonal, and matter-of-fact ways of dealing with the environment so as to limit his spontaneity. The obsessive-compulsive patient, having defense mechansims, can be less formal and may maintain a socially desirable facade. It was suggested that the use of fifty per cent or more exclusive form (F) responses in the Rorschach test should be generally interpreted as a need to avert anxiety rather than as a sign of obsessive-compulsive traits.

The use of two or more inanimate movement (m) responses was interpreted to signify the presence of obsessive-compulsive traits, since they were more frequently given by the obsessive-compulsive patient. It seemed that relatively great inner tension may prevail without the appearance of inanimate movement (m) responses. It was also concluded that the anxiety reaction patient adjusts to many situations by withdrawal and rejection while the obsessive-compulsive patient adjusts by a drive for quantity and socially acceptable behavior.

Received August 16, 1949.

References

1. Cameron, N. *Psychology of behavior disorders.* Boston: Houghton Mifflin, 1947.
2. Goldfarb, W. A definition and validation of obsessional trends in the Rorschach examination of adolescents. *Rorschach Res. Exch.* 1943, 7, 81-108.
3. Klopfer, B. and Kelly, D. *The Rorschach technique.* Yonkers: World Book Co., 1942.
4. Maslow, A. H. and Mittlemen, B. *Principles of abnormal psychology.* New York: Harper, 1941.
5. Munroe, Ruth. The inspection technique: A method of rapid evaluation of the Rorschach protocol. *Rorschach Res. Exch.*, 1944, 8, 46-70.
6. Shaffer, L. F. *The psychology of adjustment.* Boston: Houghton Mifflin, 1936.
7. Symonds, P. M. *Dynamic psychology.* New York: Appleton-Century-Crofts, 1949.
8. Thorpe, L. P. and Katz, B. *Psychology of abnormal behavior.* New York: Ronald, 1948.
9. White, R. W. *The abnormal personality.* New York: Ronald, 1948.
10. Veterans Administration Technical Bulletin, T B 10 A-78, October 1, 1947.

THE SUSCEPTIBILITY TO FALSIFICATION OF THE RORSCHACH PSYCHODIAGNOSTIC TECHNIQUE

ABRAHAM L. CARP AND ARTHUR R. SHAVZIN
UNIVERSITY OF WISCONSIN

IN 1939, L. K. Frank [4] defined projective methods as a method for the study of personality involving the presentation of a stimulus situation designed or chosen because it will mean to the subject not what the experimenter has arbitrarily decided it should mean (as in most psychological experiments using standardized stimuli in order to be "objective") but rather whatever it must mean to the personality who gives it, or who imposes upon it his private, idiosyncratic meaning and organization. The purpose of projective techniques is no different than that of the inventory-type of personality tests, that is, to provide information needed for the understanding of an individual. However, in terms of approach to personality, projectives are quite different from other personality tests. The method, in general, is to reveal the total personality, or aspects of the personality in their framework of the whole.

Because of the characteristics of projective tests, it is difficult not only to interpret projective responses quantitatively, but also to test their reliability statistically. Yet, because of their widespread use, reliability studies are of the utmost importance. Three techniques are commonly used to find the reliability of a test. They are the retest reliability, split-half reliability, and equivalent-form reliability. In general, only retest reliability can be used for projective techniques because there usually are no equivalent forms of the test, and because it is difficult, if not impossible to split the test into equivalent halves. Further, the results of a projective test depend upon the whole test. Consideration, however, must be taken of personality changes which may take place between test and retest as well as the effect of recall of previous responses on responses to the retest. (See Bell [1] for a discussion of this problem)

PROBLEM

One problem, which has ever been a nemesis to paper and pencil and other inventory-type tests is that of "faking," i.e., responding with what the subject thinks will create the best impression, rather than the true one. Some tests attempt to eliminate this problem by techniques which will show a preponderance of "lies" in the responses. The Minnesota Multiphasic [7] for instance, has an *L* score, which, if too high, indicates that the results are not reliable. Although projective techniques are believed to have a strong point in that they reduce "faking" to a minimum, there are few studies concerned with this problem.

Two studies dealing directly with faking on the Rorschach have been reported by Fosberg [2, 3]. In his first study he gave the tests to two subjects under each of four different sets of instructions. First, they were given the standard instructions, then told to "make the best impression," "the worst impression," and finally, to look specifically for certain things which the experimenter asked to be sought, e. g., "Show me all the humans in action that you can find."

After scoring and tabulating the protocols in the usual way by the Klopfer method, the tabulations of each set of four were treated statistically in the following manner. Fosberg states:

> We first proposed to test the null-hypothesis: "All four records do not vary significantly from each other." Stated another way, we can ask: "Are all four records no more than random deviations from a distribution in which the Rorschach categories are in a fixed relationship to one another?" These queries stress the probability function of our statistical treatment. Chi-square is also a test of "goodness of fit," so we can also ask: "How closely do the given frequencies fit the theoretical distribution of frequencies in the individual?" . . . . To this end we calculated the chi-square for all four tests at once.

Fosberg realized that combining Rorschach categories might be unjustified, but since the

chi-square test tends to exaggerate the discrepancy between observed and theoretical data when it is calculated with low theoretical frequencies, the only alternative was to omit categories with such low frequencies. He finally decided, after consultation with Klopfer, to take both risks, that is, combine in some cases and keep low frequencies in others. His results for the uncombined categories from which chi-square was calculated for all four tests were probabilities of .950 and .975 for the two subjects. For the combined categories for all four tests, he got probabilities of .995 and .814, respectively. On this basis he concluded that the "Rorschach test does not lend itself to manipulations based upon the subjective personal estimation of character, good, bad, or indifferent but faithfully traces the more permanent picture."

In a later study, Fosberg used a correlational technique. He used 129 subjects, both adults and children, and of both sexes, which fell into the following groups:

(1) Bernreuter "Control" Group—consisting of 37 subjects who took the Bernreuter Inventory under the usual standard instructions, then to make the best impression, the worst impression, and finally under standard conditions once more.

(2) Rorschach "Control" Group—consisting of 26 subjects reported in the literature. Each of these subjects had had two or more protocols published.

(3) Main Experimental Group — consisting of 25 males and 25 females, professors, graduate and undergraduate students of the psychology departments at New York University and Columbia University. They all took the test under the conditions mentioned for the Bernreuter Group above.

(4) Special Experimental Group—consisting of 16 subjects who took the Rorschach test as described for the main Experimental Group, except that in the last test, the Rorschach determinants were all pointed out to these subjects and they were asked to see as many as possible.

Correlations were calculated for all the groups between tests and retests. In the case of the Rorschach groups, the Rorschach Sections, "Location," "Determinants," and "Contents," were correlated separately as well as the test as a whole. Fosberg concluded that "The Rorschach withstood all the attempts at manipulation by the subjects in the two directions of making better or worse impressions of their personalities, whereas the Bernreuter Inventory could not withstand such manipulation."

Since Fosberg's studies are the only ones which are concerned with "faking" a Rorschach, it is important that his results be verified, especially so since the correlations he obtained in his second study are as high, or higher, than those obtained in reliability studies of most personality and even intelligence tests (the correlation coefficients ranged from .828 to .944). The necessity of further work is obvious. In his second study he used a larger number of subjects but used a correlation coefficient to compare the tests. He does not state, however, which type of correlation he used.[1] It is difficult to see how one correlation coefficient could be applied to the whole test since the Rorschach categories cannot be considered as class intervals. Also, in using the distribution of scores as equally related to the whole test, a difference of three where the responses in the category range from zero to three would be the same as a difference of three where the range is from 17 to 20. Further, he does not state which figures he used in making his computations. That is, it is not known from the article whether he calculated the correlation for each individual and determined the mean or whether he lumped the responses of each category together to obtain his correlation coefficients.

This study, therefore, was designed to test further the susceptibility of the Rorschach to falsification.

PROCEDURE

In the present study, the Rorschach test was given twice, three weeks apart, to each of 20 male elementary psychology students. The subjects were asked to imagine themselves in certain situations while taking the test: once, to give a "good" impression, and once, to give a "bad" impression. The directions were as follows:

> This test consists of a series of 10 cards which are designs made up out of ink blots. People see all sorts of things in these ink blots; now tell me what you see, what it might be for you, what it makes you think of. However, I want you to imagine yourself in this situation. . . .
>
> Situation I: (for bad impression) . . . . you are being drafted for the army and want very much to stay out. The results of this personality

[1]Personal communication from Dr. Fosberg states that he used the product-moment coefficient of correlation.

test will determine whether or not you will be accepted. An unfavorable report will keep you out.

Situation II: (for good impression) . . . . you are in a hospital for the mentally ill and, naturally, want to get out. You are being given this personality test to see whether or not you can be released. A favorable report will get you that release.

Now let's go through the test with these circumstances in mind. What do you see in the first card?

One half of the subjects were presented with Situation I first, the other half were presented with Situation II first. The protocols were scored and tabulated in the usual way by the Beck system.

Each category of the "good" and "bad" tests was then compared by the *t* test and correlation coefficient. Where the number of responses in a category was too small to obtain a reliable Pearson product-moment correlation, a tetrachoric correlation was calculated.

The problem of not enough responses in some categories was solved by combining certain categories which are similar. Thus, all light responses (*V, Y, FV, FY, VF, YF*) were combined, as were human responses (*H* and *Hd*), animal responses (*A* and *Ad*). Color responses were compared, both individually and as a whole. *Ct* referring to total color was calculated by weighting the color value as Beck does in calculating Experience Balance, i.e., *C*, 1.5; *CF*, 1.0; *FC*, 0.5.

The next step was to compare the whole tests, both "good" and "bad," as units, by calculating the chi-square and probability of homogeneity. Here too, categories were combined; these, in addition to those mentioned above, being the form determinants (*F+, F—*, and *F*), color (this time each color response as a unit), and a miscellaneous category consisting of all content responses other than human or animal responses. When combining the categories still did not provide enough responses for reliable calculations the category was omitted.

RESULTS

Table 1 shows the results of comparing each category by *t* test and correlation. The *t* test was used to see if a change in response, due to the conditions of this experiment, would have direction, i.e., would trying to make either a good or bad impression increase or decrease certain or all responses. It is seen that only in *Z* (organization) is the deviation from the mean significant, and that at the 5 per cent c.l. This does not mean, however, that no changes were produced. The data clearly showed the differences. But the direction taken was so diverse, among the individual subjects, that they were balanced out in the analysis. Thus it is not possible to predict in what way a subject will respond in his attempt to "fake" the results.

Five of the correlation coefficients are above .85. Three of them pertain to form (*F+, F—*, and *F*), one to time per response (T/R), and one to the number of responses per test (R). The remaining correlations ranged from .16 to .84. According to Rorschach theory, Content is most subject to manipulation, but since it is the least important of the Rorschach sections for personality interpretation, this lack of consistency is of no importance. In this study, however, the Content responses of human and animal categories had correlations of .80 and .82 respectively, while Wholes (*W*), Movement (*M*), Color-form (*CF*), and Total Color (*Ct*) had correlations of .16, .17, .21, and .33 respectively. In general, then, it must be concluded that the total responses of all the subjects for each category of the "good" test do not show a close relationship to the total responses of all the subjects for the same category of the "bad" test. These test-retest reliabilities are somewhat lower than reported for these categories when there is no attempt to distort the personality picture [5, 6].

TABLE 1
COMPARISON OF TEST CATEGORIES
(N = 20)

CATEGORY	*t*	*r*	σ	CATEGORY	*t*	*r*	σ
-W	1.00	.16	.223	Ct	1.27	.33	.356
-D	1.07	.91	.079	H+Hd	1.82	.80	.083
Dd	1.01	.84	.068	A+Ad	1.00	.82	.075
Light	1.00	.64	.135	R	1.17	.90	.044
F plus	1.06	.95	.022	Z	2.32**	.68	.123
F minus	1.43	.97*	.415	F+%	1.00	.59	.150
F	1.00	.88	.052	A %	1.00	.46	.181
M	1.00	.17*	.378	P	1.42	.55	.160
C	1.00	.65*	.391	T/1 R	1.00	.56	.157
CF	1.00	.21*	.382	T/R	1.00	.93	.031
FC	1.00	.48*	.435	S	1.53	.69*	.361

*Tetrachoric correlation coefficients.
**Significant at the 5 per cent confidence level.

To compare each subject with himself, a chi-square technique was used. It was assumed that each Rorschach category is an independent measure, thus the number of responses in each

category could be considered as the frequency. Chi-squares were calculated to determine if the two distributions from each subject were similar. Table 2 shows the results of this analysis. At the five per cent level of confidence the distribution differed significantly for three subjects and at the ten per cent level for two more subjects. It may be assumed, therefore, that under the conditions of this experiment, there were some subjects who were successful in presenting two different response patterns on the Rorschach.

TABLE 2
DEGREE OF HOMOGENEITY OF "GOOD" AND "BAD" TESTS

SUBJECT	χ^2	df	P	SUBJECT	χ^2	df	P
I	1.506	6	.95-.98	XI	13.196	9	.10-.20
II	7.840	9	.50-.70	XII	2.392	7	.90-.95
III	16.833	9	.05-.10	XIII	5.008	8	.70-.80
IV	8.935	7	.20-.30	XIV	12.604	6	.02-.05
V	7.997	7	.30-.50	XV	8.527	9	.50-.70
VI	13.834	7	.05-.10	XVI	1.678	9	.99
VII	3.866	7	.80-.90	XVII	3.315	5	.50-.70
VIII	4.539	8	.80-.90	XVIII	15.584	6	.01-.02
IX	4.651	9	.80-.90	XIX	1.702	6	.90-.95
X	6.006	5	.30-.50	XX	28.744	8	.01

DISCUSSION

The results of this study suggest that it is possible for some subjects to distort their Rorschach protocols depending on the set to make a "good" or "bad" impression. It appears that subjects' desire to "fake" must be considered. There is still, of course, the possibility that the clinician would have no difficulty in recognizing the "faked" test. Also there has been no attempt to determine whether the subjects were "successful" in altering their responses in the desired directions. But the implication that these factors need not be considered is not supported.

It should be pointed out also that the use of the chi-square technique to ascertain the similarity of the two tests is not completely justified. The test implies that each category is independent, that is, that there is, for example, no relation between *W* and *M*. This assumption is probably not tenable. The method was employed, however, to see how closely these results would approximate those obtained by Fosberg using a similar technique. A further study is in progress to determine whether by a rating and matching technique judges can differentiate between simulated tests.

SUMMARY

Twenty subjects were given the Rorschach twice under conditions designed so that they would give a "good" impression in one instance and a "bad" impression in the other. A *t* test comparison of each category in the two tests showed no group differences although there were wide individual differences in the direction of the responses in a subject's attempt to manipulate his responses. Coefficient correlations for separate scoring categories ranged from .16 to .97, suggesting that there is not a close relationship between certain categories of the "good" test and the categories of the "bad" test.

The two tests for each subject were then compared as units by means of chi-square to see whether the two distributions were similar. For four subjects the probabilities were less than .10 that the two distributions came from the same population. It is seen, therefore, that there were some subjects who were able to alter their responses.

In conclusion, the results of this study do not agree with those obtained by Fosberg, which, he states, indicate that "the Rorschach withstood all attempts at manipulation by the subjects." On the contrary, this study shows that there are some subjects who *can* manipulate their responses, who *can* vary their personality picture as reflected by the Rorschach, under instructions to make "good" or "bad" impressions.

Received August 29, 1949.

REFERENCES

1. BELL, J. E. *Projective techniques.* New York: Longmans, Green, 1948.
2. FOSBERG, I. A. Rorschach reactions under varied instructions. *Rorschach Res. Exch.*, 1938, **3**, 12-38.
3. FOSBERG, I. A. An experimental study of the reliability of the Rorschach psychodiagnostic technique. *Rorschach Res. Exch.*, 1941, **5**, 72-84.
4. FRANK, L. K. *Projective methods.* Springfield, Ill.: Charles C Thomas, 1948.
5. HERTZ, MARGUERITE R. The reliability of the Rorschach ink-blot test. *J. appl. Psychol.*, 1934, **18**, 461-477.
6. THORNTON, G. R. AND GUILFORD, J. P. The reliability and meaning of erlebnistypus scores in the Rorschach test. *J. abnorm. soc. Psychol.*, 1936, **31**, 324-330.
7. Minnesota Multiphasic Personality Inventory. New York: Psychological Corp., 1943.

A NOTE ON THE VALIDITY OF THE LEITER INTERNATIONAL PERFORMANCE SCALE

TRENT E. BESSENT
NORWALK STATE HOSPITAL, CALIFORNIA

THE Leiter International Performance Scale, hereafter referred to as the LIPS, was published in 1940 [1] as a year scale similar in design to that of the Binet but based solely upon performance items. During the war the scale was not produced, but recently an announcement has been made of its availability [2].

Procedure. Over a period of a year and a half data were gathered on twenty cases who were tested by both the Revised Stanford Binet (Form L) [3] and the LIPS (1940 edition) in the clinic of the court psychologist for Santa Barbara County, California. This sample was composed of thirteen male and seven female court cases, with fifty per cent being of Mexican ancestry. In chronological age they ranged from four years to fifty-eight years with the median age being about eleven. The middle fifty per cent ranged from about nine years to sixteen years of age. By referring to Table 1 it can be seen that the group as a whole scored rather low on the two tests with mean IQ's for both the Binet and the LIPS of about 70, and mean MA's of a little over seven years. It should also be noted that this group was more variable on the Binet than the standardizing population used on that test, the SD being over 22 instead of 16.

TABLE 1
SUMMARY OF STATISTICAL DATA ON THE REVISED STANFORD BINET AND THE LEITER INTERNATIONAL PERFORMANCE SCALE

Measure	N	Binet (L) M	Binet (L) SD	Leiter M	Leiter SD	Pearson r
IQ	20	69.85	22.61	70.55	20.65	.92±.025
MA (months)	20	88.5	36.3	86.0	33.3	.93±.021

Results. Obviously, because of the wide chronological age range and the greater variability of the group, some of the statistical findings will be spuriously large, but yet of general interest value. By referring to Table 1 the reader will find that the correlation between the IQ's of the two tests is .92. Since the IQ is an index figure, this introduces a spurious index correlation as some of the positive correlation is a result of the large CA range. However, when we eliminate the statistical problem of the index and consider the test scores in the form of the MA we find that the correlation is still a high .93. As the CA and the variability of the subjects have also influenced this correlation, an attempt was made to hold constant the CA range. In the total sample of twenty cases, six were found at the CA level of nine years. The Spearman *rho,* when comparing the MA scores of these six cases was .99.

In addition it is noted that the scores yielded by the LIPS appear to be similar in magnitude to those of the Binet. As indicated in Table 1 the mean IQ's for both tests were about 70, and statistically there were no significant differences in the means and sigmas for both the IQ and MA scores for the two tests.

Received July 21, 1949.

REFERENCES

1. LEITER, R. G. *The Leiter International Performance Scale.* Santa Barbara: Santa Barbara State Coll. Press, 1940
2. LEITER, R. G. Report on the status of the Arthur Adaptation and the 1948 Revision of the Leiter International Performance Scale. *Psychol. Serv. Cen. Bull.,* 1949, **1**, 1-7.
3. TERMAN, L. M., AND MERRILL, MAUDE A. *Measuring intelligence.* Boston: Houghton Mifflin, 1937.

NEW BOOKS AND TESTS

Books

BALES, ROBERT F. *Interaction process analysis.* Cambridge, Mass.: Addison-Wesley Press, 1950. Pp. xi + 203. $6.00.

A detailed description of the methods used by the Laboratory of Social Relations at Harvard University for the recording and analysis of the interactions in small groups. Itself a product of interdisciplinary contacts between sociology and psychology, the method of analysis is one of the most promising tools for the advancement of the study of many interpersonal areas: psychotherapy, education, social action, and politics. No recent development has shown more clearly the rewards that follow the quantification of observations that previously have been made only qualitatively and crudely.

BRYNGELSON, BRYNG, CHAPMAN, MYFANWY E., AND HANSEN, ORVETTA K. *Know yourself: A workbook for those who stutter.* (Rev. Ed.) Minneapolis: Burgess, 1950. Pp. iii + 159. $2.00.

Know yourself is a workbook planned for use by a trained speech clinician for work with those who stutter. The authors have arranged the material to give students of junior and senior high school age a working knowledge of mental responses to situations. They hope to give the student an objective attitude towards all his behavior responses, while working on the control and correction of his poor speech habits. Some of the lessons and the plan of attack will be of special interest to all who help students of that age level, such as group leaders and parents. Since many problems will be opened up, care should be taken that the lessons are presented by trained personnel. Serious difficulties might arise if problems are uncovered without the presence of a leader who is able either to give help or to direct the student to someone who can resolve his problem.—*B.M.L.*

ENGLE, T. L. *Psychology.* (Rev. Ed.) Yonkers, N. Y.: World Book Co., 1950. Pp. xi + 628.

A new edition of an elementary survey of applied psychology, for use in senior high schools and junior colleges. While much of the text is devoted to personality and personal adjustment, the selection and presentation of material often fails to achieve maximum effectiveness for the intended level of student; witness the page and a half devoted to superficial details about differences between individual, group, and multiple-choice Rorschach.

NOTE: Some reviews in this issue were prepared by the Associate Editors, who may be identified by their initials. Unsigned reviews are by the Editor.—*L.F.S.*

FELDENKRAIS, M. *Body and mature behavior.* New York: International Universities Press, 1949. Pp. viii + 167. $3.75.

The perplexing subtitle of this small book, "A Study of Anxiety, Sex, Gravitation, and Learning," may seem like a cross-out-one-wrong-word test item until the author's thesis is disclosed. Feldenkrais examines the implications of the erect human posture, and of various modes of carriage and movement, for the development of personality and for psychotherapy. The theory has its points of interest, but does not quite persuade.

HARSH, CHARLES M., AND SCHRICKEL, H. G. *Personality development and assessment.* New York: Ronald, 1950. Pp. vii + 518. $5.00.

This book is intended as a text for courses in the development and assessment of personality, or of personology, the authors having adopted Murray's term. Part I considers the problems of individuality as they arise in the developmental sequence from conception to old age. A series of chapters is concerned with the successive age levels, and in each instance the discussion covers the factors influencing personality; the major problems and mechanisms of adjustment; changes in motivation, ability and learned adjustment patterns; the significance of the phase for later developments; and characteristic individual variations. There is also a chapter on adjustments to love and marriage in early adulthood. Part II, on personality study in theory and practice, includes chapters on theory and on assessment. There is an appendix on theories of trait organization. The treatment of the various techniques for assessing and interpreting personality is necessarily sketchy, but a fair overall picture is given. The book will probably make a satisfactory text, although it displays the disadvantages as well as the advantages of a general eclecticism.—*A.R.*

HYMES, JAMES L., JR. *Teacher listen, the children speak.* New York: New York Committee on Mental Hygiene of the State Charities Aid Association, 105 E. 22 St., 1949. Pp. 44. 25¢.

An exceptionally able and attractive pamphlet, designed to help teachers understand troublesome

children. The overaggressive child, the attention-getting child, the dependent child, the withdrawing child, and the "nervous" child are crying out their need for real understanding and help.

KUBIE, LAWRENCE S. *Practical and theoretical aspects of psychoanalysis.* New York: International Universities Press, 1950. Pp. xvii + 252. $4.00.

This is a revision and expansion of a book originally published in 1936 and at that time directed toward an essentially lay audience. The present revision attempts "to present not only to laymen, but also to physicians, and to all students of psychiatry and psychology, certain theoretical as well as practical aspects of psychoanalysis, with some indication of the directions in which it may be expected to develop in the future." Your reviewer found it clear, well developed, and excellently written. It can be read with profit by all students of clinical psychology. — *W.A.H.*

LEMKAU, PAUL V. *Mental hygiene in public health.* New York: McGraw-Hill, 1949. Pp. xiv + 396. $4.50.

This book grew out of the author's experience in teaching mental hygiene to students in the School of Hygiene and Public Health at Johns Hopkins University. As such, its audience is primarily professional people who have not specialized in mental health problems. Part I, The Place of Mental Hygiene in Public Health, is a well documented argument for the necessity of integrating mental hygiene programs in the work of federal, state and local public health agencies. Part II, The Development of the Individual, is a scholarly but easy-to-read summary of personality growth and development from birth to senility. The approach here is eclectic with emphasis on the possibilities for the modification of growth in the interests of better mental health. A forty-page appendix is devoted to brief description of the major forms of psychopathology. Another appendix lists appropriate visual aids for use in mental hygiene programs. —*E.L.K.*

MEERLOO, JOOST A. M. *Patterns of panic.* New York: International Universities Press, 1950. Pp. 120. $2.00.

From his first-hand experiences in The Netherlands during World War II, the author describes and discusses the nature and causes of civilian and military panic. The approach is largely anecdotal, and provides a store of concrete illustrations perhaps unsurpassed by any other recent work. There is occasional reference to psychoanalytic formulations, but the book is not at its best when dealing with theory. The prescription for preventing panic—by preparation, information, and action—is a commonsense one, but soundly based on experience.

REIWALD, PAUL *Society and its criminals.* New York: International Universities Press, 1950. Pp. xix + 315. $4.50.

This is a psychoanalytically oriented discussion of the criminal as an outgrowth of society, and of the motivations of the social group that inflicts the punishment. The role of unconscious drives in criminal law and the need to safeguard society intelligently from crime in phantasy, and the criminal, are all analyzed in separate chapters. The book attempts to contribute to a deeper understanding of what crime means to us and how this affects our treatment of it. —*F. McK.*

SEGUIN, C. ALBERTO *Introduction to psychosomatic medicine.* New York: International Universities Press, 1950. Pp. 320. $5.00.

The comprehensiveness of this brief introduction to psychosomatic medicine results in its being overly terse at times, as it occasionally descends to what is little more than outline form. Consequently the selection and treatment of material is somewhat uneven. There are six pages devoted to the anatomy of the hypothalamus and only three to its physiology; while sixty pages (of the total three hundred) are allotted to the taking of the psychosomatic clinical history. This unevenness detracts from the value of the book as an all-around text, but it is essentially sound, and could be used with value as supplementary reading at the introductory level. Even the experienced worker will find much that is stimulating, as in the author's brief but penetrating discussion of the concepts of "disease" and "cure." —*W.A.H.*

SMART, MOLLIE STEVENS *Babe in a house.* New York: Scribner's, 1950. Pp. 212. $2.75.

A really good popular book about babies and their families, that psychologists can recommend safely to their layman friends. Mainly by easy-reading case studies, it gives many of the insights that a normal but perplexed parent needs. Mrs. Smart not only recognizes, but communicates effectively, that the same "method" will not be best for all babies, nor for all parents. She avoids the middle-of-the-road confusions that often try to pass for moderation, and presents a sound reasonableness that is based on a good knowledge of modern psychological research and theory, and a sensitive understanding of direct experience.

SONNEMANN, ULRICH *Handwriting analysis as a psychodiagnostic tool.* New York: Grune & Stratton, 1950. Pp. x + 276. $5.00.

Why has graphology so long remained a rejected stepchild of psychology in America, while receiving more recognition in Europe? A thoughtful study of Sonnemann's book perhaps gives an answer. Graphology, at least the "systematic" graphology of Klages that is expounded here, is content to describe the character of a person in terms of the character of his handwriting. The test and the referent are one and the same, whereas the empirical culture of American psychology demands an external referent. Illustrations are easily found in Sonnemann. In the chapter "Twelve Lines from

Joan,'" illustrating the application of his analysis, are many phrases such as these: "it seems to caress and cultivate rather than neglect the letter forms. . .," "the quantitative extension of the movements is well measured; it never gets involved in 'crippling' over-scrupulous detail. . . ." American psychology wants to know something more. Does Joan "caress and cultivate" her children and her friends? Does she "never get involved in overscrupulous detail" in her home and in her work? Such questions remain unanswered. It will be no solution to regress to atomistic letter-measuring in order to be objective and quantitative, for, as Sonnemann rightly says, that approach has been unrewarding. However, newer global approaches being developed for the validation of projective techniques may well prove applicable to graphology also. Meanwhile, clinical psychology may withhold judgment, neither rejecting graphology as quackery, nor espousing it too warmly, until the further evidence becomes available.

STEINER, M. E. *The psychologist in industry.* Springfield, Ill.: Charles C Thomas, 1949. Pp. vii + 107. $2.00.

A lecture-in-print on the roles of the industrial psychologist in selecting and placing workers, in working with them on the job, and in counseling them. For so small a volume, the summaries of research are surprisingly full. The digest of the applications of the Rorschach to industrial situations is especially complete, although sometimes uncritical. The newer responsibility of the psychologist as a personal counselor in industry is recognized adequately.

STECKLE, LYNDE C. *Problems of human adjustment.* New York: Harper, 1949. Pp. x + 351. $3.00.

The author states that in this work he attempts "to introduce man to himself." Through generous use of examples he develops the theory that "the health of an individual is determined by the reaction between inner necessity and external demands." Application of his thesis is discussed in terms of the individual's adjustment to five areas of adult life: sexual adjustment, courtship and marriage, work, religion and old age. Though the author states clearly that adult behavior is learned behavior, he places greater emphasis on the effect in adult life of adaptive and nonadaptive behavior than on conditions in early life creating these responses. The style and subject matter of this book are personalized and directed toward stimulation of self-appraisal.

TREDGOLD, R. F. *Human relations in modern industry.* New York: International Universities Press, 1950. Pp. 192. $2.50.

This volume consists, in the main, of a series of lectures delivered at a course in human relations in industry at Roffey Park Rehabilitation Center, London, England. The lectures were originally designed as stimuli to discussion for the participants of the course—psychiatrists, industrial physicians, nurses, personnel officers, psychiatric social workers, a union leader, and others. Because of the mixed audience for whom these lectures were designed, it may be unfair to assess the contents at face value, but the fact remains that psychologists and trained personnel workers will find the treatment too elementary. Psychiatric principles are stated in oversimplified terms, and principles of "human relations" are in terms of nonpsychiatric capsules. Those not oriented in modern psychology may find the mixture of psychological fragments, inspirational examples, and homespun quotations interesting. —*M.K.*

1950 Directory of Vocational Counseling Services. Ethical Practices Committee, National Vocational Guidance Association. St. Louis 5, Mo.: Ethical Practices Com., Box 64, Washington Univ., 1950. Pp. 98. $1.00.

"Sooner or later every profession that wins public acceptance faces the problem of policing its own practitioners, and helping the lay public to identify the honest workman," states the president of the N. V.G.A. in his foreword to this important pamphlet. Culminating continuous efforts since 1921, the *Directory* indicates that the profession of vocational guidance has indeed reached a high degree of maturity. The minimum standards for vocational guidance service are described explicitly in an introductory section, in terms of recognition, personnel, procedures, advertising, and fees. The body of the pamphlet gives a geographical listing of 131 agencies and individual practitioners, each of which was investigated by questionnaire, by references, and by a visit. Listing is voluntary, and only agencies which apply are studied. For each agency, the sponsorship, services, clientele, fees, method of application, director, staff, and hours of service are stated. Psychologists throughout the country will find the pamphlet an invaluable guide to comunity resources.

TESTS

AMMONS, ROBERT B., BUTLER, MARGARET N. AND HERZIG, SAM A. *Vocational Apperception Test.* High school-college-adult. Individual test. 18 plates. (10 for men, 8 for women), with manual, pp. 7 ($7.25). Louisville, Ky.: R. B. Ammons, Univ. of Louisville, 1949.

An experimental edition of a set of pictures, each representing one occupation. Stories, obtained and analyzed in a manner similar to the TAT, are intended to evoke expressions of occupational preference, reason for entering an occupation, personal needs and mechanisms, and outcomes. The manual gives only meager data concerning results.

LANGMUIR, CHARLES R. *Oral Directions Test.* Revised manual, pp. 8. New York: Psychological Corp., 1950.

The revised manual for this phonograph-administered ability test gives additional normative data for children, veterans, and industrial workers, and includes some findings on the validity of the method for industrial selection.

JOURNAL OF CONSULTING PSYCHOLOGY
AVAILABLE BACK ISSUES

The issues listed below are available for sale

YEAR	VOLUME	JAN FEB	MAR APR	MAY JUN	JUL AUG	SEP OCT	NOV DEC	PRICE PER NUMBER	PRICE PER VOLUME
1937	1	1	2	3	4	5	6	$.60	$3.00
1938	2	1	2	3	4	5	6	$.60	$3.00
1939	3	1	2	3	4	5	6	$.60	$3.00
1940	4	1	2	3	4	5	6	$.60	$3.00
1941	5	1	2	3	4	5	6	$.60	$3.00
1942	6	1	–	3	–	–	6	$.60	$1.80
1943	7	–	2	3	4	5	6	$.60	$3.00
1944	8	1	–	3	4	5	6	$.60	$3.00
1945	9	–	2	3	4	5	6	$.60	$3.00
1946	10	1	2	3	4	5	6	$.60	$3.00
1947	11	–	–	3	4	5	6	$.60	$2.40
1948	12	1	2	3	4	5	6	$1.00	$5.00
1949	13	1	2	3	4	5	6	$1.00	$5.00
1950	14			By Subscription, $5.00				$1.00	

Price for Volumes 1 through 13 $41.20

Table based on inventory of January 1, 1950

The Journal of Consulting Psychology is of particular interest to clinical psychologists, psychiatrists, school psychologists, and persons who are engaged in counseling and guidance work.

The price per volume of the Journal of Consulting Psychology was $3.00 between 1937 and 1947. It is now $5.00 per volume and $1.00 per issue. For foreign orders, $.25 per volume should be added. Address orders to:

AMERICAN PSYCHOLOGICAL ASSOCIATION
1515 Massachusetts Ave. N. W.
Washington 5, D. C.

PSYCHOLOGICAL BULLETIN

July
No. 4
1947

THE PRESENT STATUS OF PSYCHOTHERAPEUTIC COUNSELING

BY

WILLIAM U. SNYDER

This popular review is the July, 1947 issue of the Psychological Bulletin

Pp. 90 $1.25

AMERICAN PSYCHOLOGICAL ASSOCIATION
1515 MASSACHUSETTS AVENUE, N. W.
WASHINGTON 5, D. C.